Relativity Theory and Astrophysics

I. Relativity and Cosmology

Volume 8
Lectures in Applied Mathematics

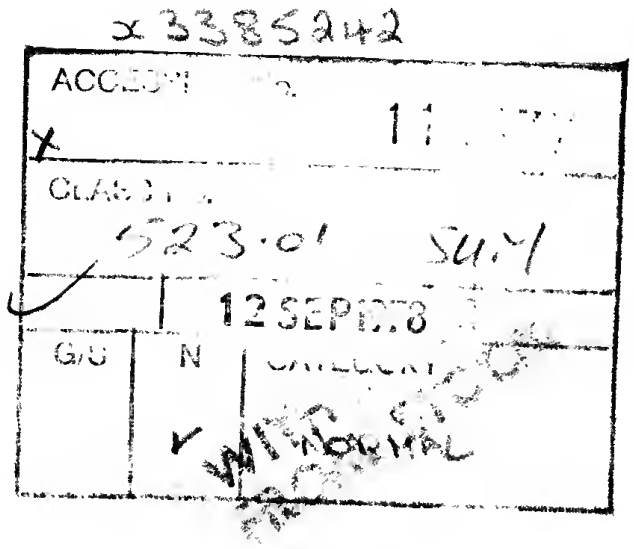
Volume 8
Lectures in Applied Mathematics

RELATIVITY THEORY AND ASTROPHYSICS

1. RELATIVITY AND COSMOLOGY

Jürgen Ehlers, EDITOR
The University of Texas

1967
American Mathematical Society, Providence, Rhode Island



The Summer Seminar was conducted, and the praceedings prepared in part, by the American Mathematical Society under the fallawing cantracts and grants:

Air Farce Office af Scientific Research under Grant AF-834-65
Atamic Energy Cammissian under Contract AT(30)1364-2
Natianal Aeronautics and Space Administratian under Grant NsG 358
Natianal Science Faundatian under NSF Grant GE-7790
Office af Naval Research under Cantract Nanr(G)00052-65.

All rights reserved except thase granted ta the United States Gavernment, otherwise, this baak, or parts thereof, may nat be repraduced in any farm withaut permissian af the publishers.

Library af Cangress Catalag Card Number 62-21481
International Standard Baak Number 0-8218-1108-0
Capyright © 1967 by the American Mathematical Society
Reprinted with corrections 1972
Printed in the United States af America

Foreword

This volume is the first of a series of three which contain the Proceedings of the Fourth Summer Seminar on Applied Mathematics, arranged by the American Mathematical Society and held at Cornell University from July 26 to August 20, 1965.

The purpose of the Seminar was primarily to acquaint graduate students and recent recipients of the Ph.D. degree with the state of knowledge and current problems in Relativity and Astrophysics and thus to stimulate research in these subjects.

Because the participants could not be expected to be familiar with basic information in both the fields of relativity and astrophysics, five series of basic lectures were given. These consisted of ten lectures on the Theory of Relativity by A. Schild, four lectures on Theoretical Cosmology by E. Schücking, seven lectures on Galactic Structure and Galactic Dynamics by L. Woltjer, seven lectures on Stellar Structure by E. E. Salpeter, and six lectures on Stability Problems by S. Chandrasekhar. In addition, one or more lectures were given on Experimental Tests of General Relativity, Relativistic Hydrodynamics, Gravitational Collapse, Gravitational Radiation, Observational Cosmology, Cooperative Phenomena in Stellar Dynamics, Spiral Structure of Galaxies, and Cosmic Rays; a survey of the lectures is given in the Introduction contained in this volume.

The program of the Seminar was organized by a committee consisting of the following members:

S. Chandrasekhar
C. C. Lin
C. W. Misner
A. Schild
A. H. Taub (Chairman).

Thanks are due to the five government agencies which supplied financial support, as acknowledged on the copyright page; to the staff of the American Mathematical Society for administrating the Seminar, particularly to Dr. Gordon L. Walker, Executive Director of the Society, who contributed greatly to the planning and functioning of the Seminar; to the members of the organizing committee listed above, especially its chairman; to Dr. William Smith, Director of Summer Session and Extramural Courses of Cornell University, and to Professor Martin Harwit, Department of Astronomy, for handling the housing and for the provision of other facilities for the Seminar and staff. The devotion of Mrs. M. L. Leigh to the discharging of her duties in the day to day operation of the Seminar office deserves special mention.

JÜRGEN EHLERS, EDITOR

THE UNIVERSITY OF TEXAS
AUSTIN, TEXAS

Contents

INTRODUCTION

I. Relativity

LECTURES ON GENERAL RELATIVITY THEORY 1
Alfred Schild

COMPARISON OF THEORY AND OBSERVATION IN GENERAL RELATIVITY 105
L. I. Schiff

EXPERIMENTAL TESTS OF GENERAL RELATIVITY. RADAR 117
Freeman J. Dyson

GRAVITATIONAL WAVES 129
R. K. Sachs

MEASUREMENTS OF SPACE TIME CURVATURE AND SEARCH FOR GRAVITATIONAL RADIATION 141
J. Weber

CONSERVED QUANTITIES AND CONFORMAL STRUCTURE IN GENERAL RELATIVITY 147
Roger Penrose

TAUB-NUT SPACE AS A COUNTEREXAMPLE TO ALMOST ANYTHING 160
C. W. Misner

RELATIVISTIC HYDRODYNAMICS 170
A. H. Taub

GRAVITATIONAL COLLAPSE 194
G. C. McVittie

DRAGGING OF INERTIAL FRAMES BY ROTATING MASSES 200
Jeffrey M. Cohen

II. Cosmology

LECTURES ON QUASI-STELLAR OBJECTS 203
Maarten Schmidt

COSMOLOGY 218
E. L. Schücking

A UNIFIED APPROACH TO COSMOLOGY 237
David Layzer

JEANS' FORMULA FOR GRAVITATIONAL INSTABILITY 263
W. B. Bonnor

MICROWAVE RADIATION FROM THE BIG BANG 274
P. J. E. Peebles

INDEX 285

Introduction

Gravitation is considered to be of fundamental importance to physics and astrophysics for the following two reasons:

(1) All neutral, structureless test particles move equally in the neighborhood of massive bodies (Eötvös-Dicke Experiment), hence the gravitational field cannot be separated unambiguously from the inertial field (Einstein's principle of equivalence). Therefore and because of the approximate local validity of the special theory of relativity, space-time is represented in Einstein's general theory of relativity as a pseudo-Riemannian manifold whose curvature measures the inhomogeneity of the gravitational field. Almost all physical laws require for their mathematical formulation a metric and a linear connection; these structures are physically defined by the behavior of particles[1] which, unavoidably, are subject to gravitational forces. Consequently the gravitational field influences, in principle, all physical processes.

(2) Among all known basic interactions gravitation is uniquely distinguished by its long range and accumulative character; the latter property, due to the positivity of mass, discerns it from elec-

[1] This statement has been stressed above all by H. Weyl in his classic book, *Space, Time, Matter*; it has recently been illustrated particularly clearly and emphatically by J. A. Wheeler.

tromagnetism. Gravitation, therefore, is believed to govern (sometimes in conjunction with large-scale magnetic fields, perhaps) the structure and motion of macroscopic matter distributions such as stars, star clusters, galaxies, clusters of galaxies, and the system of "all" the latter, the main object of cosmological studies.

The following lectures, presented at the American Mathematical Society Summer Seminar on Relativity and Astrophysics which was held from July 26 to August 20, 1965 at Cornell University, deal with very different topics and methods of approach; what unites them is that they are all concerned with gravitation. The contributions have been arranged, independently of the sequence in which the lectures were delivered, such as to indicate the systematic connection of the various topics with each other. The intention of this Introduction is to provide a bird's-eye view of the whole set of lectures and thus to make apparent the interrelationships between its parts. For reference purposes the third contribution to volume 9, e.g., will be called lecture (9.3).

The theories of gravitation used in the following lectures are Einstein's and Newton's. The principal concepts and laws of Einstein's theory are presented in A. Schild's lectures (8.1). Newton's theory, used throughout volume 9 and in most of E. E. Salpeter's article (10.1) on stellar structure, is critically reviewed in §C of E. L. Schücking's contribution (8.12) on cosmology. Arguments which indicate that Newton's theory is a good approximation to Einstein's under certain assumptions are sketched in §4 of Schild's lectures.

The mathematical models used for the (macroscopic) description of matter are those of fluid dynamics and of kinetic theory. A. H. Taub describes, after some remarks on the (special-) relativistic kinetic theory of a simple gas, the relativistic theory of ideal fluids in lectures (8.8). The post-Newtonian approximation to the equations of hydrodynamics with gravitational interaction is developed by S. Chandrasekhar in (10.2); his equations provide a quantitative estimate of the domain and degree of validity of the Newtonian theory considered as an approximation to Einstein's.[2] The basic ideas and equations of kinetic theory as applied to dilute

[2] Observe, however, the remark to be made in connection with (10.1).

systems of gravitationally interacting particles are reviewed by L. Woltjer in lecture (9.1).

In view of the central position of the Liouville equation in galactic structure theory (volume 9) and the importance of the radiative transfer equation in the theory of stellar structure it may be worth pointing out that one can begin with the general-relativistic Boltzmann equation for systems of particles of arbitrary rest masses and with arbitrary collisions (due to short range forces). The specialization to massive particles without collisions then gives the general-relativistic Liouville equation which might be useful for cosmology, and the weak-field limit of the latter is the nonrelativistic Vlasov equation. The specialization to photons interacting with a material medium gives, far quasi-equilibrium distributions, the equation of radiative transfer. Also, the hydrodynamical equations can be obtained as moment equations, as is well known in the nonrelativistic case and is indicated by A. H. Taub in the relativistic case.

Within the general framework indicated, essentially five groups of problems are treated in the three volumes on Relativity and Astrophysics:

A. Empirical tests of general relativity in the laboratory and in the solar system,
B. Theoretical developments in general relativity,
C. Cosmology,
D. Galactic structure,
E. Stellar structure.

We proceed to describe in turn these groups of lectures.

A. The classical three tests of Einstein's theory are treated in Schild's §9, and further proposed tests are discussed by L. I. Schiff, F. J. Dyson and J. Weber. The dynamical theory underlying the proposed gyroscope experiment discussed by Schiff (8.2), an experiment that would permit testing the tensor character of gravitation, is essentially contained in the elegant derivation of the equations of motion of a spinning spherical test particle by Schild [see lecture (8.1), Equations (6.46), (6.47)].

The reader may find it illuminating to establish the relation between Schild's and Schiff's treatments, and he might ask himself which generalization yields, for a nonspherical particle, the "Newtonian" precession.

Very promising for the experimental investigation of quasi-stationary gravitational effects in the solar system is the rapid development of radar techniques discussed by F. J. Dyson (8.3).

The prospects of the skilful and persistent attempts of J. Weber [lecture (8.5)] concerning the measurement of gravitational radiation seem still to be rather uncertain, in spite of the remarkable improvement in sensitivity of the receiver and power of the emitter that have been achieved.

B. The theory of gravitational radiation is treated by A. Schild [lecture (8.1), §§7 and 10] and by R. K. Sachs [lecture (8.4)]; some supplements are contained in Weber's lectures (8.5). One of the latest and most exciting offsprings of this theory, the theory of ten quantities which are conserved even if the source radiates gravitationally, is introduced in (8.6) by one of its creators, R. Penrose, who also explains a powerful conformal technique for handling asymptotic properties of radiation fields.

C. W. Misner shows in lecture (8.7) by means of an example which peculiar geometrical properties occur in relativistic space-time manifolds; further investigations along these lines would seem to be necessary in order that some insight be obtained into the global implications of Einstein's field equations, mathematically as well as in regard to the physical interpretation of global models.

The general statements on relativistic hydrodynamics and interior solutions of Einstein's equations made by A. H. Taub in lectures (8.8) are supplemented by G. C. McVittie in lecture (8.9) by some examples; the models of collapsing, spherically symmetrical masses are related to some stability considerations of C. Hunter [lecture (9.6)] and should be contrasted with the more realistic models presented by M. M. May and R. H. White in (10.3).

J. M. Cohen in (8.10) shows how local inertial frames are dragged along by rotating, hollow bodies, a "Machian" effect of general relativity.

C. The cosmological part of volume 8 begins with a discussion of what are probably the most distant signposts in space-time that have been observed so far, the quasi-stellar objects. M. Schmidt (8.11) describes their observed characteristics and discusses in particular the interpretations of their large red-shifts.

E. L. Schücking (8.12) reviews the isotropic models of the universe according to Newton's and Einstein's theories of gravitation, stressing the great similarity between these two approaches. Combining statistical assumptions with Einstein's theory D. Layzer (8.13) offers ideas which might help to understand the lumpy

structure (clustering) of the observed matter distribution in the universe; in his treatment Newton's theory appears as a local approximation.

The old and vexing question how one can understand the formation of galaxies is taken up again by W. B. Bonnor (8.14) who presents a derivation of Jeans's criterion for gravitational instability by determining the time development of spherically symmetrical "small" perturbations superimposed on a (Newtonian) isotropic model universe, and finds that gravitational instability alone can hardly account for galaxy formation.

As shown by P. J. E. Peebles in lecture (8.15), some observational evidence for the most basic cosmogonic assumption underlying the Robertson-Walker type relativistic models, namely, the existence and uniformity of an ultrahot initial plasma, seems to be provided now by the microwave background radiation observed since 1965.

Another (controversial) question of cosmological significance, that of the density of cosmic radiation outside of galaxies, is briefly discussed by J. Linsley in lecture (9.8), and also touched in M. Burbidge's lecture (9.7).

D. The basic facts on galaxies are reported in the first and third sections of L. Woltjer's lectures (9.1). A semi-empirical, dynamical theory of a stationary, purely stellar galaxy, based on the collisionless Boltzmann—and Poisson equations, is outlined in §II; several appendices complete quantitatively the material discussed. The most conspicuous structural feature of many galaxies, a spiral pattern, is the main subject of C. C. Lin's lectures (9.2). Using the coupled equations for the motions of stars and gas (formulated in Appendix I), it is shown how the spirals might be understood as density waves propagating through the medium formed by stars and gas.

Since the distribution function of a stationary stellar system is constant along stellar orbits (Liouville's theorem), it is important to know which isolating integrals are associated with which (average, gravitational) potentials. This problem which is related to the classification of orbits is treated by G. Contopoulos in (9.3).

The theory of spherical, gas-free star clusters is described concisely by J. R. King in lecture (9.4). The different roles of mixing, relaxation and evolution (escape) for establishing the actual distribution are elucidated, and the construction of models is indicated.

The general, formal theory of stability for solutions of the Liouville-Poisson equations and related topics—the propagation of small disturbances of various kinds, the time scales involved, etc.—are discussed in lectures (9.5) by D. Lynden-Bell. The diversity of both the astrophysical phenomena involved as well as the mathematical tools used—Fourier-Laplace transformations, analytic continuation, spectral theory of self-adjoint linear transformations—and the many open problems make these lectures particularly stimulating and will hopefully induce further research.

A comparison of the derivations of Jeans's criterion given by Bonnor and by Lynden-Bell respectively suggests that a common generalization of both versions—namely when the background material is isotropically expanding and the matter is treated as a statistical ensemble of particles rather than as a continuum—should be possible, at the relativistic and at the nonrelativistic level.

The important cosmogonic problems of star-formation by condensation, for the treatment of which stability theory is one of the main tools, is next considered. C. Hunter sketches in lecture (9.6) how a galactic HI cloud might fragment into protostars, and which roles are played in this complex process by gravitational, thermal, rotational and magnetic phenomena.

Two observationally oriented lectures form the end of volume 9: E. M. Burbidge's authoritative, detailed survey (9.7) of the observed properties of radio galaxies and the problems they pose to the theoreticians, and J. Linsley's report (9.8) on composition, directional distribution, spectral properties and the possible origin of cosmic radiation.

E. The basic observational data on stars and the structure theory of "normal" nonrotating stars in quasi-equilibrium states are presented in the first two sections of the lectures (10.1) by E. E. Salpeter. In the later sections more exotic objects—extremely massive stars, white dwarfs, neutron stars, and supernovae—and the associated physical phenomena of radiation pressure, electron degeneracy, nucleon degeneracy, strong gravitational fields, etc., are discussed; here links are established with general relativity. As pointed out in §3 of (10.1), general relativistic corrections are required not only in cases with strong gravitational fields, i.e., where the ratio gravitational radius/geometrical radius of some body is close to unity; they may influence the stability of a body already if its binding energy is sufficiently small in comparison with

the Newtonian potential energy. This remark suggests the systematic investigation of the influence of post-Newtonian terms on the dynamics, particularly the stability, of gaseous masses, a complex of problems to which the major part of lectures (10.2) by S. Chandrasekhar is devoted. Using a Coulomb-type gauge-condition[3] in conjunction with a weak-field approximation, Chandrasekhar derives the post-Newtonian equations of hydrodynamics; the conservation laws for mass, energy, linear momentum, angular momentum; and the tensor and scalar virial theorems. These formulae are then used to study the stability of spherical gaseous masses and the relativistic corrections to the equilibrium properties of rotating fluids. These lectures, therefore, provide a continuation of lectures (8.1) and (8.8), and complete or extend various stability considerations in lectures (9.6) and (10.1).

The last two lectures are devoted to the collapse of spherical masses according to general relativity. M. M. May and R. H. White present in (10.3) the results of numerical calculations on collapsing perfect fluid masses, the primary aim being to complement dynamically the results on a maximum stable mass for equilibrium configurations, and to study the motion of fluid spheres which have approximately the critical mass. C. W. Misner in (10.4) sets up the equations for spherical collapse with inclusion of (purely radial) radiation, describes the development of "trapped surfaces" in the interior parts of a collapsing body, and introduces a new coordinate system which is better suited to the computation of effects which can be noticed by an outside observer rather than to the dynamics of the unobservable, inner parts.

It emerges from the lectures presented in these three volumes that there are many more open problems than solved ones in the domain of gravitational phenomena, and that many approaches are tried to transfer problems from the first category to the second one. If these lectures should stimulate such researches, and if they succeed in increasing the interest of scientists and mathematicians in the fields of relativity and astrophysics, they will have served their purpose.

J. EHLERS, Editor

[3] One equation, in contradistinction to Fock et al., who use the four harmonic, Lorentz type gauge conditions.

I. RELATIVITY

Alfred Schild

Lectures on General Relativity Theory[1]

1. **Review of special relativity theory (kinematics).** In classical physics (roughly pre-1900), the laws of mechanics cannot distinguish between different inertial frames, but the laws of electrodynamics pick out one preferred inertial frame, the ether or the frame of absolute rest.

Empirical evidence shows that none of the laws of physics can distinguish between different inertial frames: that there is no ether. This is the principle of relativity. The evidence also shows that in each inertial frame Maxwell's equations are valid, and their consequence, the isotropic propagation of light with a constant speed c (we choose units so that $c = 1$). The theory of special relativity is based on these facts. It implies that the transformations from one inertial frame to another are the Lorentz transformations, rather than the classical Galilean transformations. It implies that

[1] Supported in part by the Aerospace Laboratories, Office of Aerospace Research, United States Air Force. These lectures were also given in the Spring of 1966 at the Neils Bohr Institute and the Nordic Institute for Theoretical Atomic Physics (Copenhagen), where the author was a visiting professor. He wishes to thank Professor Christian Møller and his colleagues for a stimulating year spent in a hospitable institute and a friendly country.

we must change our classical concepts of space and time, and that these merge into a single four-dimensional world.

Minkowski Space-Time M_4. This is a flat 4-space with signature $- - - +$. In rectangular Cartesian coordinates, the metric is

$$\begin{aligned} ds^2 &= -dx^2 - dy^2 - dz^2 + dt^2 \\ &= \eta_{mn} dx^m dx^n, \end{aligned} \tag{1.1}$$

where lower case Latin suffixes range and sum over 1, 2, 3, 4, where $x^m = (x, y, z, t)$ and

$$\eta_{mn} = \eta^{mn} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{1.2}$$

Rectangular Cartesian coordinates correspond physically to inertial frames. The Lorentz transformations are the transformations from one rectangular Cartesian coordinate system to another. A law of physics automatically satisfies the principle of relativity if it is written as a tensor equation in M_4.

The *null* (or light) *cone* of an event 0 (a point $x_0{}^m$ of M_4) consists of all events x^m connected with 0 by a light signal, travelling with speed 1. It is given by the invariant equation $\Delta s^2 = 0$, where $\Delta s^2 = \eta_{mn} \Delta x^m \Delta x^n$ and $\Delta x^m = x^m - x_0{}^m$. The null cone divides the points of space-time into the following regions (Figure 1):

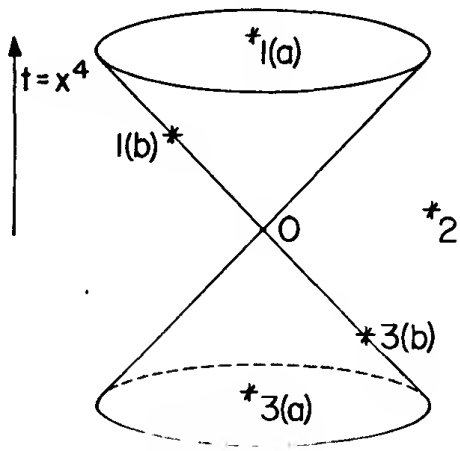


FIGURE 1

1.Absolute future of 0: (a) interior of future null cone: $\Delta s^2 > 0$, $\Delta x^4 > 0$; (b) future null cone: $\Delta s^2 = 0$, $\Delta x^4 > 0$.

2. Relative "present" of 0: exterior of null cone: $\Delta s^2 < 0$.

3. Absolute past of 0: (a) interior of past null cone: $\Delta s^2 > 0$, $\Delta x^4 < 0$; (b) past null cone: $\Delta s^2 = 0$, $\Delta x^4 < 0$.

The motion of a particle is represented by a curve in space-time, the *world line* of the particle. A material particle has a time-like world line, i.e., a curve for which $ds^2 > 0$ or, equivalently, a curve which lies inside the null cone of each of its points (Figure 2). This corresponds to a particle speed $v < 1$ in any inertial frame. The arc length

$$s = \int_{A(\Gamma)}^{B} ds \tag{1.3}$$

between events A and B of a time-like world line is the physical time measured by the particle itself, or by a clock moving with the particle. It is called the proper time.

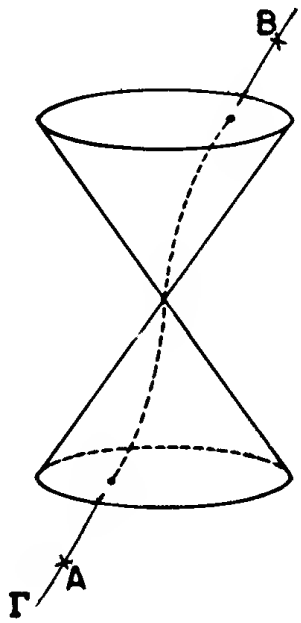


FIGURE 2

In general, two curves joining the same two points A and B will have different lengths: $s_1 \neq s_2$ (Figure 3). This gives the famous

clock or twin effect of relativity theory: Two identical clocks (twins), showing the same time (age) at event A, are separated and have different motions; if they are reunited at an event B, they will in general show different times (ages).[2]

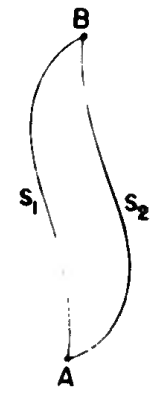


FIGURE 3

A free material particle, under no forces, has a straight time-like world line. The motion of a light pulse is represented by a straight null line (Figure 4), along which $ds^2 = 0$, corresponding to a speed $v = 1$ in any inertial frame.

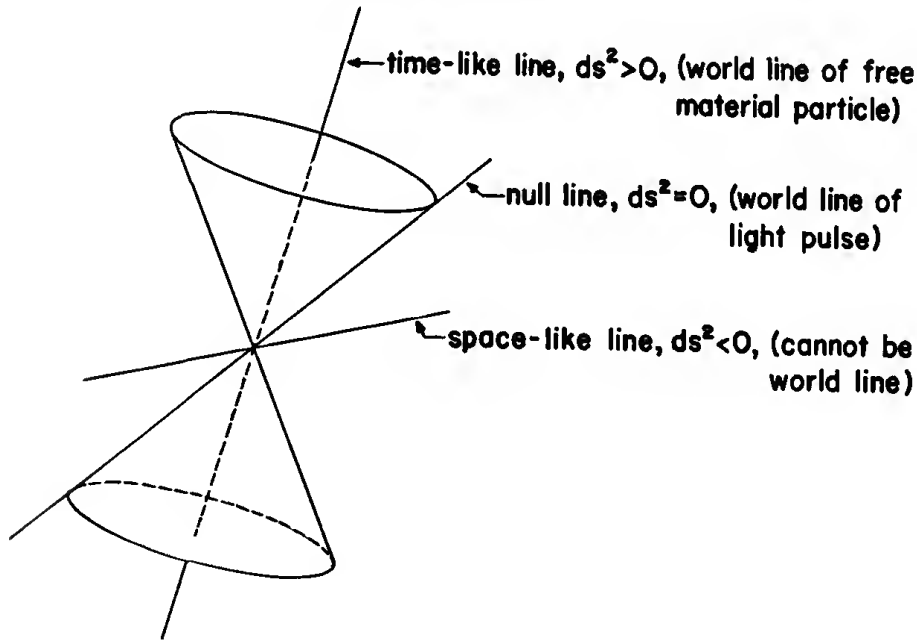


FIGURE 4

[2] A. Schild, Amer. Math. Monthly 66 (1959), 1.

Lorentz Transformations. They are the transformations which leave η_{mn} invariant:

$$\eta_{mn} = \frac{\partial x'^r}{\partial x^m}\frac{\partial x'^s}{\partial x^n}\eta_{rs}. \tag{1.4}$$

Mathematically they are the transformations from one rectangular Cartesian coordinate system in M_4 to another, physically the transformations from one inertial frame to another. These transformations are easily shown to be linear (exercise):

$$\begin{aligned} x'^r &= L^r{}_s x^s + L^r, \\ \eta_{mn} &= L^r{}_m L^s{}_n \eta_{rs}. \end{aligned} \tag{1.5}$$

This condition is equivalent to (exercise) $\eta^{mn} = L^m{}_r L^n{}_s \eta^{rs}$, and implies

$$L = \det(L^r{}_s) = \pm 1. \tag{1.6}$$

Also $-(L^4{}_1)^2 - (L^4{}_2)^2 - (L^4{}_3)^2 + (L^4{}_4)^2 = 1$, so that

$$L^4{}_4 \geqq 1 \quad \text{or} \quad L^4{}_4 \leqq -1. \tag{1.7}$$

The transformations with $L^r = 0$ are called homogeneous; with $L = +1$ proper, with $L = -1$ improper; with $L^4{}_4 \geqq 1$ orthochronous (they take absolute future into absolute future), with $L^4{}_4 \leqq -1$ antichronous (they interchange absolute future and absolute past). The homogeneous proper orthochronous Lorentz transformations form a group $\mathscr{L}_+^{\uparrow}$.

Exercise. Under a Lorentz transformation (1.5), a tensor transforms according to

$$T'^{r\cdots}{}_{s\cdots} = L^r{}_m \cdots L_s{}^n \cdots T^{m\cdots}{}_{n\cdots}, \tag{1.8}$$

where $L_s{}^n = \eta_{sr}\eta^{nm}L^r{}_m$.

Any time-like straight line can be chosen as the t'-axis of an inertial frame.

Exercise. Show that if the first of a pair of twins remains at rest in an inertial frame, while the second twin travels away and returns, then the second twin will be younger than the first.

Because Lorentz transformations are linear, straight lines will look like straight lines in diagrams in Minkowski space-time, parallel lines will look like parallel lines. We can trust our Euclidean

intuition as far as projective properties (points, straight lines) and affine properties (parallels) are concerned. These are sufficient to compare lengths in parallel directions. Thus, if in Figure 5, $ABDC$, $ABED$ and $ABFE$ are parallelograms, then CF is 3 times the length of the parallel segment AB.

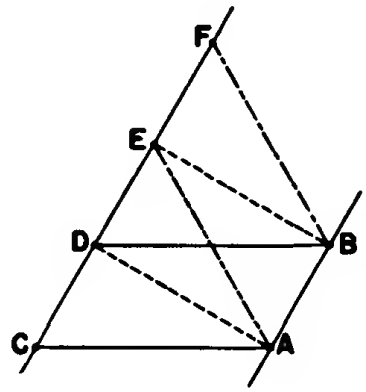


FIGURE 5

Where metrical properties are concerned, we can no longer trust our Euclidean intuition. This is because of the different signature of Minkowski space-time. Right angles do not look like right angles, equal lengths in different directions do not always look equal. Thus, in Figure 6, $t'0x'$ is a right angle and the four points shown are at unit distance from 0. The figure illustrates the simplest Lorentz transformation between two inertial frames in uniform relative motion along a common x-direction in space, with y- and z-axes parallel.

Spinors. Spinors provide an important representation of Lorentz transformations. Spin space is a complex two-dimensional space, where vectors are transformed by unimodular linear transformations:

$$T'^A = \lambda^A{}_B T^B, \tag{1.9}$$

$$\lambda = \det(\lambda^A{}_B) = 1. \tag{1.10}$$

Capital Latin suffixes range and sum over 1, 2. The complex conjugate of T^A is denoted by $\overline{T}^{A'}$ and transforms according to

$$\overline{T}'^{A'} = \overline{\lambda}^{A'}{}_{B'} \overline{T}^{B'} \tag{1.11}$$

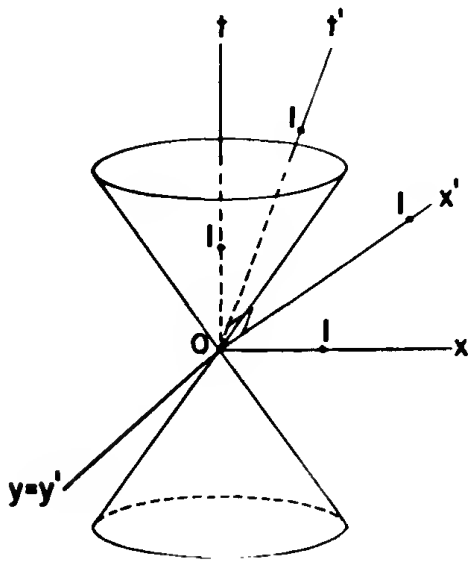


FIGURE 6

where $\overline{\lambda}^{A'}{}_{B'}$ is the complex conjugate of $\lambda^{A}{}_{B}$. Suffixes are raised and lowered by means of a skew symmetric spin metric

$$\epsilon_{AB} = \epsilon^{AB} = \epsilon_{A'B'} = \epsilon^{A'B'} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{1.12}$$

according to the rules $T^{A} = \epsilon^{AB}T_{B}$, $T_{A} = T^{B}\epsilon_{BA}$,

$$\overline{T}^{A'} = \epsilon^{A'B'}\overline{T}_{B'}, \qquad \overline{T}_{A'} = \overline{T}^{B'}\epsilon_{B'A'}.$$

A general spin tensor, or spinor, transforms as follows:

$$\begin{aligned} &T'^{A\cdots}{}_{B\cdots}{}^{C'\cdots}{}_{D'\cdots} \\ &\quad = \lambda^{A}{}_{P}\cdots\Lambda_{B}{}^{Q}\cdots\overline{\lambda}^{C'}{}_{R'}\cdots\overline{\Lambda}_{D'}{}^{S'}\cdots T^{P\cdots}{}_{Q\cdots}{}^{R'\cdots}{}_{S'\cdots}, \end{aligned} \tag{1.13}$$

where $\Lambda_{A}{}^{P}\lambda^{B}{}_{P} = \delta^{B}_{A}$ and $\overline{\Lambda}_{A'}{}^{P'}\overline{\lambda}^{B'}{}_{P'} = \delta^{B'}_{A'}$.

The spin metric is invariant under spin transformations, i.e., $\epsilon'_{AB} = \epsilon_{AB}$, etc. (exercise). We adopt the convention that the complex conjugate of a spinor is denoted by

$$\overline{T^{A\cdots}{}_{B\cdots}{}^{C'\cdots}{}_{D'\cdots}} = \overline{T}^{A'\cdots}{}_{B'\cdots}{}^{C\cdots}{}_{D\cdots}. \tag{1.14}$$

The important formula[3]

$$\epsilon_{AB}\epsilon_{CD} + \epsilon_{AC}\epsilon_{DB} + \epsilon_{AD}\epsilon_{BC} \equiv 3\epsilon_{A[B}\epsilon_{CD]} = 0 \tag{1.15}$$

follows from the fact that the left-hand side is completely skew symmetric in the three suffixes B, C, D, which can take on only two values, 1 and 2.

Exercise. Show that

$$T_A{}^A = -T^A{}_A, \quad \epsilon_A{}^B = \delta_A^B, \tag{1.16}$$

$$T_{AB} = T_{(AB)} + \frac{1}{2}\epsilon_{AB}T_C{}^C, \tag{1.17}$$

$$A_A = \lambda B_A \Leftrightarrow A_A B^A = 0. \tag{1.18}$$

A correspondence $T^a \leftrightarrow T^{AB'}$ between a real world vector T^a and a Hermitian spin tensor $T^{AB'} = \overline{T}^{B'A}$ is established by the relations

$$\begin{aligned} T^1 &= (T^{12'} + T^{21'})/2^{1/2}, \quad T^2 = i(T^{12'} - T^{21'})/2^{1/2}, \\ T^3 &= (T^{11'} - T^{22'})/2^{1/2}, \quad T^4 = (T^{11'} + T^{22'})/2^{1/2}, \end{aligned} \tag{1.19}$$

$$\begin{aligned} T^{11'} &= (T^3 + T^4)/2^{1/2}, \quad T^{12'} = (T^1 - iT^2)/2^{1/2}, \\ T^{21'} &= (T^1 + iT^2)/2^{1/2}, \quad T^{22'} = (-T^3 + T^4)/2^{1/2}. \end{aligned} \tag{1.20}$$

From these equations we find

$$T_a T^a = 2\det(T^{AB'}). \tag{1.21}$$

Since spin transformations are unimodular (1.10), they leave $\det(T^{AB'})$ invariant. It follows that a spin transformation induces a homogeneous Lorentz transformation, and since the former can be obtained continuously from the identity ($\lambda^A{}_B = \delta_B^A$), the latter must be proper and orthochronous. Clearly $\pm\lambda^A{}_B$ correspond to the same Lorentz transformation.

The spin transformations

---

[3] Symmetrization is denoted by round brackets, e.g., $F_{(ab)} = (F_{ab} + F_{ba})/2$; skew symmetrization by square brackets, e.g.,

$$F_{[abc]} = (F_{abc} + F_{bca} + F_{cab} - F_{bac} - F_{acb} - F_{cba})/6.$$

The same notation applies to spinor or other suffixes.

$$\lambda^A{}_B = \begin{pmatrix} e^{i\theta/2} & 0 \\ 0 & e^{-i\theta/2} \end{pmatrix}, \quad \lambda^A{}_B = \begin{pmatrix} \cos\tfrac{1}{2}\theta & \sin\tfrac{1}{2}\theta \\ -\sin\tfrac{1}{2}\theta & \cos\tfrac{1}{2}\theta \end{pmatrix},$$

(1.22)

$$\lambda^A{}_B = \begin{pmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{pmatrix}$$

with θ, λ real, correspond respectively to rotations through angles θ in the xy- and zx-planes and to a Lorentz transformation in the zt-plane for a velocity $v = (1-\lambda^4)/(1+\lambda^4)$ between the inertial frames (exercise). Every homogeneous, proper, orthochronous Lorentz transformation can be obtained by a sequence of these special transformations.

We therefore have the one to two correspondence

$$L^r{}_s \leftrightarrow \pm\lambda^A{}_B \tag{1.23}$$

between the group $\mathscr{L}_+{}^\uparrow$ and the group Λ of spin transformations. Near the identity ($L^r{}_s - \delta^r{}_s$ and $\lambda^A{}_B - \delta^A_B$ small), the correspondence is one to one. Also Λ is easily seen to be simply connected. Therefore Λ is isomorphic to the covering group of $\mathscr{L}_+{}^\uparrow$, and $\mathscr{L}_+{}^\uparrow$ is doubly connected.

Equations (1.19) and (1.20) are a special case of the general correspondence between world tensors and (special) spinors, $T_a{}^{b\cdots}_{\cdots} \leftrightarrow T_{AC'BD'\cdots}$, given by

$$T_a{}^{b\cdots}_{\cdots} = \sigma_a{}^{AC'}\sigma^{bBD'}\cdots T_{AC'BD'\cdots}$$

$$\Leftrightarrow T_{AC'BD'\cdots} = \sigma^a{}_{AC'}\sigma_{bBD'}\cdots T_a{}^{b\cdots}_{\cdots}, \tag{1.24}$$

where $\sigma_a{}^{AB'}$ are the Hermitian matrices

$$\sigma_1{}^{AB'} = \frac{1}{2^{1/2}}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2{}^{AB'} = \frac{1}{2^{1/2}}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix},$$

(1.25)

$$\sigma_3{}^{AB'} = \frac{1}{2^{1/2}}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \sigma_4{}^{AB'} = \frac{1}{2^{1/2}}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

The lower case suffix on $\sigma_a{}^{AB'}$ is raised and lowered by the Minkowski metric η_{ab}, the capital suffixes by the spin metric ϵ_{AB} or $\epsilon_{A'B'}$.

Exercise. Show that

$$\sigma_a{}^{AB'} = \bar{\sigma}_a{}^{B'A}, \tag{1.26}$$

$$\sigma_a{}^A{}_{C'}\sigma_b{}^{BC'} + \sigma_b{}^A{}_{C'}\sigma_a{}^{BC'} = \eta_{ab}\epsilon^{AB}, \tag{1.27}$$

$$\eta_{ab}\sigma^a{}_{AC'}\sigma^b{}_{BD'} = \epsilon_{AB}\epsilon_{C'D'}. \tag{1.28}$$

Exercise. If

$$E_{abcd} \leftrightarrow E_{AE'BF'CG'DH'} = i(\epsilon_{AC}\epsilon_{BD}\epsilon_{E'H'}\epsilon_{F'G'} - \epsilon_{AD}\epsilon_{BC}\epsilon_{E'G'}\epsilon_{F'H'}), \tag{1.29}$$

show that $E_{abcd} = \epsilon_{abcd}$, the completely skew symmetric Levi-Civita symbol defined by $\epsilon_{abcd} = \epsilon_{[abcd]}$, $\epsilon_{1234} = 1$. [Hint: Show that $E_{abcd} = -E_{bacd} = -E_{abdc} = -E_{acbd}$, $E_{1234} = 1$].

2. **Review of special relativity theory (dynamics).**

Particle Mechanics. The *world velocity* of a material particle, with world line $x^m(s)$, is the unit tangent vector (Figure 7)

$$u^m = \frac{dx^m}{ds}, \qquad u^m u_m = 1. \tag{2.1}$$

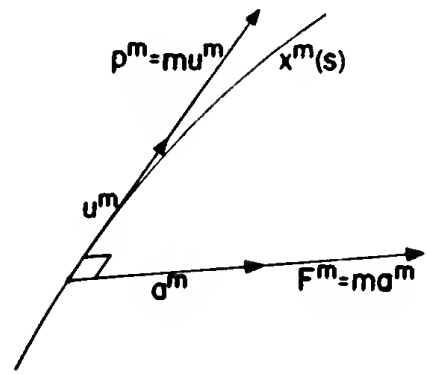


FIGURE 7

The *world acceleration* is

$$a^m = \frac{du^m}{ds} = \frac{d^2x^m}{ds^2}, \qquad a^m u_m = 0.$$

Geometrically, the direction of a^m is that of the principal normal, the magnitude $a = (-a^m a_m)^{1/2}$ is the curvature. Physically, a is the magnitude of intrinsic acceleration, which would determine the apparent weight felt by the pilot of a rocket with the given world line.

In terms of the Newtonian or relative velocity v^μ (Greek suffixes will range and sum over 1, 2, 3, unless indicated otherwise), we have

$$u^m = \frac{1}{(1-v^2)^{1/2}}(v^\mu; 1), \qquad v^2 = v^\mu v^\mu. \tag{2.2}$$

In the instantaneous rest frame of the particle,

$$u^m = (0,0,0,1), \qquad a^m = (d^2x^\mu/dt^2, 0). \tag{2.3}$$

With a material particle is associated an invariant, its *mass* (or rest mass) m. The *world momentum* is

$$p^m = mu^m = \frac{m}{(1-v^2)^{1/2}}(v^\mu, 1). \tag{2.4}$$

The fourth component

$$E = p^4 = m + \frac{1}{2}mv^2 + \frac{3}{8}mv^4 + \cdots \tag{2.5}$$

is the energy of the particle, which includes the rest energy $E_0 = m$, the classical kinetic energy $\frac{1}{2}mv^2$ and relativistic correction terms. The equations of motion of a particle are

$$\frac{dp^m}{ds} = ma^m = F^m, \qquad F^m u_m = 0. \tag{2.6}$$

F^m is called the *world force*. Only three of these equations are independent.

With a photon we associate the *wave vector*

$$k^m = \nu(n^\mu, 1), \qquad k^m k_m = 0, \tag{2.7}$$

where ν is the frequency and n^μ the unit vector ($n^\mu n^\mu = 1$) in space in the direction of propagation. Since the phase $k_m(x^m - x_0^m)$ is an invariant, k^m is a world vector. Planck's law $E = h\nu$ is generalized to the vector equation

$$p^m = hk^m = h\nu(n^\mu, 1), \qquad p^m p_m = 0, \tag{2.8}$$

the null vector p^m being the world momentum of the photon. More generally, with a particle of zero rest mass (such as a neutrino) a null vector is associated as world momentum.

When particles collide, the total world momentum is unchanged (Figure 8):

$$\Sigma p^m = \bar{\Sigma}\bar{p}^m. \tag{2.9}$$

The number of particles before and after the collision need not be the same. If they are, and if each particle retains its rest mass, then the collision is called elastic.

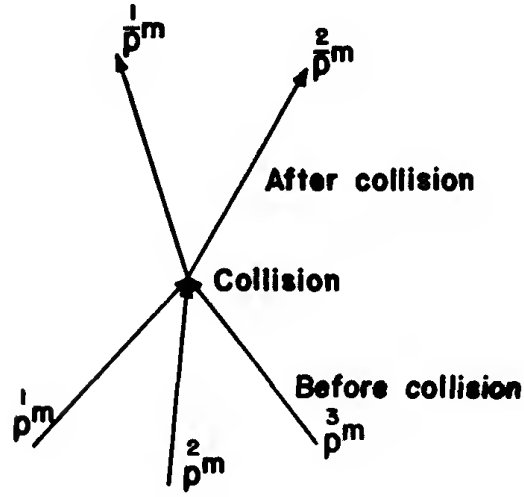


FIGURE 8

Exercise. Draw a space-time diagram to convince yourself that an electron and a positron cannot annihilate into a single photon.

Electromagnetism. The electromagnetic field (outside matter) is characterized by a skew symmetric world tensor

$$F_{mn} = -F_{nm}, \tag{2.10}$$

its sources by the *charge-current world vector* J^m. *Maxwell's equations* are

$$F_{[mn,r]} = 0, \tag{2.11}$$

$$F^{mn}{}_{,n} = 4\pi J^m. \tag{2.12}$$

Here and throughout, suffixes following a comma denote partial differentiation, e.g., $F_{mn,r} = \partial F_{mn}/\partial x^r$.

Introducing the dual $A^*{}_{mn}$ of a skew symmetric tensor A_{mn} by

$$A^*{}_{mn} = \frac{1}{2}\epsilon_{mnrs}A^{rs}, \tag{2.13}$$

the first set of Maxwell's equations (2.11) can be written in the equivalent form

$$F^{*mn}{}_{,n} = 0. \tag{2.14}$$

The field equations (2.11)-(2.12) reduce to the classical equations by making the identification $E^\mu = F_{\mu 4}$, $H^\mu = (F_{23}, F_{31}, F_{12})$, $J^m = (I^\mu, \sigma)$, where E^μ and H^μ are the electric and magnetic vectors in 3-space, σ is the charge density and $I^\mu = \sigma v^\mu$ the current density.

The first set of Maxwell's equations is equivalent to the existence of *electromagnetic potentials* A_m:

$$F_{mn} = A_{m,n} - A_{n,m}. \tag{2.15}$$

The world vector $A^m = (A^\mu, \phi)$ combines the usual vector potential A^μ and scalar potential ϕ in 3-space.

The world force which the external field exerts on a point charge e, moving with world velocity u^m, is

$$F^m = eu^n F_n{}^m. \tag{2.16}$$

This clearly satisfies the important requirement $F^m u_m = 0$. The three independent components F^μ reduce in the nonrelativistic limit ($v^2 \ll 1$) to the usual Lorentz force. For a continuous distribution of charges and currents, the world force is replaced by the world vector

$$f^m = J^n F_n{}^m; \tag{2.17}$$

in the nonrelativistic limit f^μ is the Lorentz force per unit volume in 3-space.

Green's Theorem and Conservation of Charge. Green's theorem states

$$\int_{\tau(4)} T_{,m} d\tau_{(4)} = \oint_{\tau(3)} T d\tau_{(3)m}. \tag{2.18}$$

Here $\tau_{(4)}$ is a 4-region bounded by the closed 3-dimensional hypersurface $\tau_{(3)}$; $d\tau_{(4)} = \epsilon_{mnrs} d_{(1)}x^m d_{(2)}x^n d_{(3)}x^r d_{(4)}x^s$ is the 4-extension of a 4-cell (hyper-parallelepiped) of $\tau_{(4)}$ spanned by the independent infinitesimal displacements $d_{(1)}x^m$, $d_{(2)}x^m$, $d_{(3)}x^m$, $d_{(4)}x^m$; the order of these vectors (orientation of the 4-cell) is chosen so that $d\tau_{(4)} > 0$; $d\tau_{(3)m} = \epsilon_{pqrm} d_{(1)}x^p d_{(2)}x^q d_{(3)}x^r$ is the 3-extension of a 3-cell (parallelepiped) spanned by the independent infinitesimal displacements $d_{(1)}x^m$, $d_{(2)}x^m$, $d_{(3)}x^m$ lying in $\tau_{(3)}$; the orientation is chosen so that $d\tau_{(3)m}\lambda^m > 0$ for any vector λ^m pointing outward from $\tau_{(4)}$ (Figure 9). This is a nonmetrical form of Green's theorem and holds quite generally. We shall apply it later to the curved space-time of general relativity theory.

From Maxwell's equations (2.12), we immediately have

$$J^m{}_{,m} = 0, \tag{2.19}$$

the law of conservation of charge in differential form (exercise:

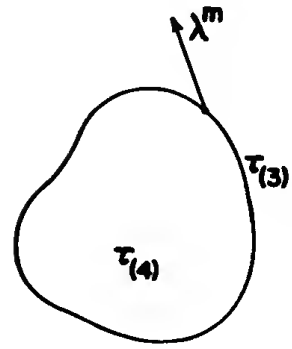


FIGURE 9

write (2.19) in 3-dimensional notation). Green's theorem gives the conservation law in integral form:

$$\oint_{\tau(3)} J^m d\tau_{(3)m} = 0 \tag{2.20}$$

for any closed $\tau_{(3)}$.

Consider an isolated system of charges and currents, a time-like world tube outside of which $J^m = 0$ (Figure 10). Apply the integral conservation law (2.20) to the hypersurface which bounds the 4-region $\tau_{(4)}$ shown in the figure. The only surviving terms are

$$\int_{\bar{\tau}(3)} J^m d\tau_{(3)m} - \int_{\tau(3)} J^m d\tau_{(3)m} = 0, \tag{2.21}$$

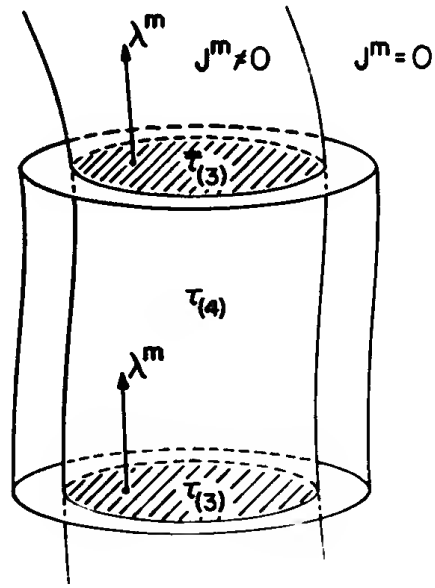


FIGURE 10

where on the two sections $\tau_{(3)}$ and $\bar{\tau}_{(3)}$ of the tube, the orientation of $d\tau_{(3)m}$ has been chosen so that $d\tau_{(3)m}\lambda^m > 0$ for λ^m pointing into the future part of the tube. Thus

$$e = \int_{\tau_{(3)}} J^m d\tau_{(3)m} \tag{2.22}$$

is independent of the particular section $\tau_{(3)}$ over which the integral is evaluated. Making the special choice $t =$ constant for $\tau_{(3)}$, and $d_{(1)}x^m = (dx, 0, 0, 0)$, $d_{(2)}x^m = (0, dy, 0, 0)$, $d_{(3)}x^m = (0, 0, dz, 0)$ for the vectors spanning $d\tau_{(3)m}$, we obtain

$$e = \int_{t=\text{const.}} \sigma dxdydz. \tag{2.23}$$

Hence e is the *total charge* of the isolated system. Our results show that e is constant in time and Lorentz invariant.

Continuum Mechanics. Continuous matter in hydrodynamic flow is described by a symmetric world tensor, the *material stress energy tensor*

$$P^{mn} = P^{nm}, \tag{2.24}$$

which satisfies the differential equations

$$P^{mn}{}_{,n} = f^m. \tag{2.25}$$

In terms of classical variables in 3-space: $P^{44} = \rho$, the mass density; $P^{\mu 4} = \rho v^\mu$, where v^μ is the macroscopic velocity; $P^{\mu\nu} = \rho v^\mu v^\nu + p^{\mu\nu}$, where $p^{\mu\nu} = p^{\nu\mu}$ is the stress. The world vector f^m is the world force density; in the nonrelativistic limit $f^4 = 0$, and f^μ is the body force per unit volume.

In the nonrelativistic limit, Equations (2.25) reduce to the usual equations of hydrodynamics, the first three equations ($m = \mu$) become the equations of motion of the fluid, the fourth equation becomes the equation of continuity (conservation of mass).

Let u^m be the unit time-like eigenvector of P^{mn}:

$$P^{mn}u_n = \tilde{\rho}u^m. \tag{2.26}$$

We call u^m the local world velocity of the fluid, and $\tilde{\rho}$ the proper mass density. The proper frame $\tilde{S}$ of the fluid at an event is ob-

tained by taking the t-axis in the direction of u^m. In the proper frame, $\tilde{\rho}$ is the mass density, $\tilde{v}^\mu = 0$ and $\tilde{P}^{\mu 4} = 0$.

A fluid is said to be *perfect* if the stress in the proper frame is diagonal: $\tilde{p}^{\mu\nu} = \tilde{p}\delta^\mu{}_\nu$; $\tilde{p}$ is called the proper pressure. In a general inertial frame, the stress energy tensor of a perfect fluid is easily seen to have the form

$$P^{mn} = (\tilde{\rho} + \tilde{p})\, u^m u^n - \tilde{p}\eta^{mn}, \quad u^m u_m = 1. \tag{2.27}$$

When $\tilde{p} = 0$, we say that we have *dust* or *incoherent matter*; in this case

$$P^{mn} = \tilde{\rho} u^m u^n, \qquad u^m u_m = 1. \tag{2.28}$$

The term f^m in the right-hand side of the hydrodynamic equations (2.25) represents the body forces exerted on the fluid by some external field. Such external forces can be made internal by enlarging the fluid system to include the field, and by including contributions of the field to the stress energy tensor. We shall illustrate this for the case where the fluid is charged and where the body forces are exerted by an electromagnetic field and are given by Equation (2.17).

Define the *Maxwell stress energy tensor* by

$$E^{mn} = \frac{1}{4\pi}\left[- F^m{}_a F^{na} + \frac{1}{4}\eta^{mn} F^{ab} F_{ab} \right]. \tag{2.29}$$

If we differentiate and use Maxwell's equations, we obtain (exercise):

$$E^{mn}{}_{,n} = - J^n F_n{}^m = - f^m. \tag{2.30}$$

Thus the *total stress energy tensor* of fluid and field,

$$T^{mn} = P^{mn} + E^{mn}, \tag{2.31}$$

satisfies

$$T^{mn} = T^{nm}, \tag{2.32}$$

$$T^{mn}{}_{,n} = 0. \tag{2.33}$$

Equations (2.33) are the differential conservation laws of energy (or mass) and momentum. The total energy density $T^{44} = \rho$ is assumed to be positive in every inertial frame. Thus

(2.34) $$T^{mn}\lambda_m\lambda_n > 0$$

for all time-like vectors λ_m, provided $T^{mn} \neq 0$.

The *angular momentum density* about an event a^m is defined by the world tensor

(2.35) $$s_{(a)}{}^{mnr} = -s_{(a)}{}^{nmr} = 2(x^{[m} - a^{[m}) T^{n]r}.$$

Clearly,

(2.36) $$s_{(a)}{}^{mnr} = s_{(0)}{}^{mnr} - 2a^{[m}T^{n]r}.$$

As a consequence of Equations (2.32)-(2.33), we have the differential conservation laws

(2.37) $$s_{(a)}{}^{mnr}{}_{,r} = 0.$$

Isolated System; Mass Center. Consider an isolated system, a time-like world tube outside of which $T^{mn} = 0$, and therefore also $s_{(a)}{}^{mnr} = 0$. The same argument which led to the expression (2.22) for charge, now shows that the differential conservation laws (2.33) and (2.37) lead to integral conservation laws. The expressions

(2.38) $$P^m = \int_{\tau(3)} T^{mn} d\tau_{(3)n},$$

(2.39) $$S_{(a)}{}^{mn} = -S_{(a)}{}^{nm} = \int_{\tau(3)} s_{(a)}{}^{mnr} d\tau_{(3)r}$$

are tensors associated with the tube as a whole, and independent of the particular section $\tau_{(3)}$ over which the integrals are evaluated. We call P^m the *world momentum* vector of the isolated system, and $S_{(a)}{}^{mn}$ the *angular momentum* about the event a^m. We assume that P^m is time-like. We also have

(2.40) $$S_{(a)}{}^{mn} = S_{(0)}{}^{mn} - 2a^{[m}P^{n]}.$$

Three of the six components of $S_{(a)}{}^{mn}$ correspond to the classical angular momentum, the other three components serve to determine the center of mass of the isolated system. This will now be shown.

Consider the points $\bar{x}^m$ which satisfy

(2.41) $$0 = S_{(\bar{x})}{}^{mn}P_n = S_{(0)}{}^{mn}P_n - 2\bar{x}^{[m}P^{n]}P_n.$$

This condition determines the straight line parallel to P^m, given by

$$\bar{x}^m = \frac{S_{(0)}{}^{mn} P_n}{P^a P_a} + \mu P^m, \tag{2.42}$$

where μ is a parameter. Let us work in the inertial frame which has this line as t-axis. In this frame

$$\begin{aligned} P^\mu &= \int_{t=\text{const}} T^{\mu 4} dxdydz = 0, \\ P^4 &= \int_{t=\text{const}} T^{44} dxdydz = \int_{t=\text{const}} \rho\, dxdydz, \\ \bar{x}^\mu &= 0, \qquad S_{(\bar{x})}{}^{\mu\nu} = \int_{t=\text{const}} 2x^{[\mu} T^{\nu]4} dxdydz, \\ S_{(\bar{x})}{}^{\mu 4} &= \int_{t=\text{const}} x^\mu T^{44} dxdydz = \int_{t=\text{const}} x^\mu \rho\, dxdydz = 0. \end{aligned} \tag{2.43}$$

The last equation shows that the line (2.42) is a center of mass. We call it the *relativistic center of mass* (Figure 11).

Less specialized mass centers can be defined.[4] Consider a time-like curve $\bar{x}^m(s)$ such that at any of its points $\bar{x}^m$ we have (Figure 11)

$$S_{(\bar{x})}{}^{mn} \frac{d\bar{x}_n}{ds} = 0. \tag{2.44}$$

In the instantaneous rest frame at that point, i.e., in the frame where $\bar{x}^m = 0$ and $d\bar{x}^m/ds = (0,0,0,1)$, this becomes

$$\int_{\tau(3):\, t=\bar{x}^4=0} \rho\, x^\mu dxdydz = 0. \tag{2.45}$$

We call the curve $\bar{x}^m(s)$ a *local center of mass.*

We now return to the proper frame of the system which has the relativistic mass center as t-axis, so that $P^\mu = 0$, $P^4 = P$, $S_{(\bar{x})}{}^{\mu 4} = 0$. By rotating the space axes, we can also arrange it that $S_{(\bar{x})}{}^{23} = S_{(\bar{x})}{}^{31} = 0$, $S_{(\bar{x})}{}^{12} = S$. In this frame, the motion of a local mass center is given by (exercise):

$$\frac{d\bar{\bar{x}}}{d\bar{\bar{t}}} = \frac{P}{S}\bar{\bar{y}}, \quad \frac{d\bar{\bar{y}}}{d\bar{\bar{t}}} = -\frac{P}{S}\bar{\bar{x}}, \quad \bar{\bar{z}} = 0. \tag{2.46}$$

[4] C. Møller, Communications of the Dublin Institute for Advanced Studies, Series A, No. 5 (1949); Ann. Inst. Henri Poincaré, (5) **11** (1950), 251; M. H. L. Pryce, Proc. Roy. Soc. London A **195** (1948), 62.

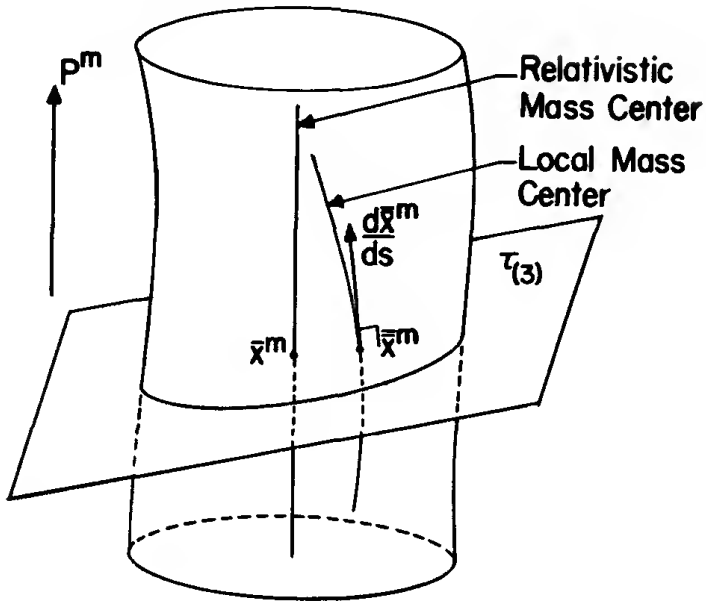


FIGURE 11

Thus, in the proper frame, the local center of mass moves with constant speed $\bar{\bar{v}}$ on a circle, whose plane is orthogonal to the proper spin $S_{(\bar{x})}{}^{\mu} = (S_{(\bar{x})}{}^{23}, S_{(\bar{x})}{}^{31}, S_{(\bar{x})}{}^{12})$, and whose radius $\bar{\bar{r}}$ is

$$\bar{\bar{r}} = \frac{S}{P}\bar{\bar{v}}. \tag{2.47}$$

The sense of the circular motion is opposite to that of the proper spin (Figure 12).

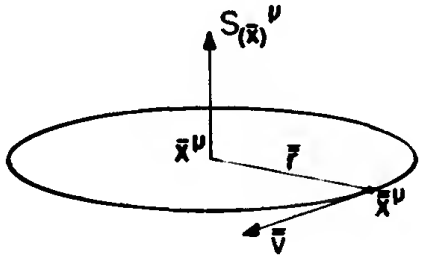


FIGURE 12

Exercise. Let $x^m(s)$ denote a local center of mass, $u^m = dx^m/ds$, $S^{mn} = S_{(x)}{}^{mn}$ the angular momentum about the local mass center, and $m = P^m u_m$. Show that

$$\frac{d}{ds}\left(mu^m - \frac{du_n}{ds}S^{mn}\right) = 0, \tag{2.48}$$

$$\frac{dS^{mn}}{ds} - u^m u_a \frac{dS^{an}}{ds} - u^n u_a \frac{dS^{ma}}{ds} = 0, \tag{2.49}$$

$$S^{mn} u_n = 0.$$

A time-like world tube is called convex if each of its plane space-like sections is convex. The convex hull of an isolated system is the smallest world tube which is convex and contains the system.

From Equation (2.45), and the fact that ρ is nonnegative, it follows that every local center of mass lies within the convex hull of the system. Since, in Equation (2.47), we can approach the maximum radius

$$\bar{r}_{\text{Max}} = S/P \tag{2.50}$$

by letting $\bar{v} \to 1$, we obtain the important result: An isolated system with proper spin S and proper mass P cannot be too small; more precisely, in the proper frame of the system, its convex hull contains a disk of radius S/P in the plane perpendicular to the proper spin.

3. **The principle of equivalence.** The theory of special relativity shows that there is no preferred inertial frame or ether. The theory of general relativity shows that there are no inertial frames at all.

Gravitation has a unique feature. The gravitational charge (or mass) of a body equals its inertial mass; all bodies accelerate equally in a gravitational field. This has now been verified by the Dicke-Eötvös experiments[5] with accuracy of one part in 10^{11}. It is therefore impossible to distinguish by observation of falling bodies between the uniform acceleration of a noninertial frame and a uniform gravitational field in an inertial frame. Einstein's principle of equivalence generalizes this fact and asserts that no physical experiment whatever can distinguish between the two possibilities. The principle of equivalence applies to a region of space which is sufficiently small to treat the gravitational field as homogeneous (Einstein's elevator).

A small region of space sweeps out a thin tube in space-time. We therefore state the principle of equivalence as follows:

Any physical experiment within a thin time-like space-time tube in a gravitational field gives identical results, to the first

[5] P. G. Roll, R. Krotkov and R. H. Dicke, Ann. Phys. **26** (1964), 442.

order in the small lateral dimensions of the tube, as the corresponding experiment in a suitable thin tube in Minkowski space-time free of gravitational fields. We adopt units so that the constant of gravitation $G = 1$. Then "thin" means that the lateral dimensions of the tube are small compared to the reciprocal of the gradient of the gravitational force field.

If we introduce curvilinear coordinates in Minkowski space-time, the metric

$$ds^2 = -dx^2 - dy^2 - dz^2 + dt^2 \tag{3.1}$$

becomes

$$ds^2 = g_{mn}dx^m dx^n, \tag{3.2}$$

where the metric tensor g_{mn} is a function of the coordinates x^m. The converse is not true. If we assign functions $g_{mn}(x)$ we have, in general, a curved Riemannian space, and there exists no rectangular Cartesian coordinate system where the metric has the form (3.1) throughout the space. Now rectangular Cartesian coordinates in space-time correspond physically to inertial frames, and the principle of equivalence indicates that such frames do not exist.

This suggests the basic hypothesis of Einstein's theory of general relativity: space-time is a Riemannian 4-space with signature $- - - +$. The physical interpretation of ds is the same as in special relativity. Along the time-like world line of a material particle, ds is the element of physical time measured by the particle itself, the element of proper time. The propagation of light is given by $ds = 0$.

As we shall see in more detail later on, the role of the Newtonian gravitational potential V is taken over by the ten components g_{mn} of the metric tensor.

Given a point Q in a Riemannian space-time, we can always[6] introduce coordinates so that

$$g_{mn} = \eta_{mn}, \quad g_{mn,r} = 0 \tag{3.3}$$

at the point Q. These coordinates are said to be *geodesic* at Q. We can go further. Given a time-like curve Γ, we can always introduce

[6] Statements such as this are not meant to be global; they apply to a finite, but sufficiently small region of space-time.

Fermi Coordinates of the First Kind.[7] They are defined by the property that Equations (3.3) hold at all points of the curve Γ. These coordinates may be constructed as follows (Figure 13):

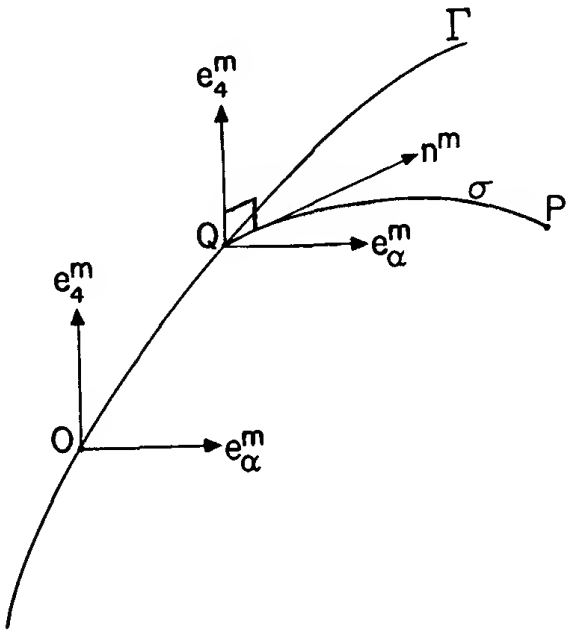


FIGURE 13

At a point 0 of Γ choose an orthonormal tetrad of vectors $e_\alpha{}^m$, $e_4{}^m$:

$$e_\alpha{}^m e_{\beta m} = -\delta_{\alpha\beta}, \quad e_\alpha{}^m e_{4m} = 0, \quad e_4{}^m e_{4m} = 1. \tag{3.4}$$

Here the first suffix on the vectors $e_\alpha{}^m$, $e_4{}^m$ is a tetrad label, the second suffix is a tensor suffix, Greek suffixes range and sum over 1, 2, 3. Propagate the four tetrad vectors parallelly along Γ, so that they are defined at all points of the curve and satisfy Equations (3.4) along Γ. At a point Q of Γ, let n^m be a general space-like unit vector lying in the 3-space spanned by $e_\alpha{}^m$:

$$n^m = n^\alpha e_\alpha{}^m, \qquad n^\alpha n^\alpha = 1. \tag{3.5}$$

Draw the geodesic through Q with n^m as tangent. Let P be a point on this geodesic, the arc length QP being $\sigma = \int_Q^P |ds|$. The choice of Q, n^α and σ uniquely determines all points P sufficiently close to Γ.

For a point Q on Γ, we define the Fermi coordinates

[7] E. Fermi, Rend. Acc. Lincei **21** (1922), 21, 51; T. Levi-Civita, Math. Ann. **97** (1927), 291.

$$(3.6) \qquad (x'^{\mu})_Q = -\int_{O(\Gamma)}^{Q} e_{\mu r}\,dx^r, \qquad (x'^4)_Q = \int_{O(\Gamma)}^{Q} e_{4r}\,dx^r,$$

and for a general point P

$$(3.7) \qquad (x'^{\mu})_P = (x'^{\mu})_Q + n^{\mu}\sigma, \qquad (x'^4)_P = (x'^4)_Q.$$

It is left as an exercise to show that the new coordinates (after dropping primes) do in fact satisfy the conditions (3.3) all along the curve Γ.

When Γ is a geodesic, we can simplify the coordinates further. We choose $e_4{}^m$ initially tangent to Γ at 0. It then remains tangential all along Γ. We call such coordinates *geodesic Fermi coordinates*. In these coordinates the curve Γ has the equations $x^{\mu} = 0$.

The Laws of Physics in General Relativity Theory. When there is a real gravitational field and space-time is curved, Fermi coordinates of the first kind are the closest analogue to inertial frames in special relativity. If the time-like curve Γ in curved space-time has equations $x^m(s)$ in Fermi coordinates, we can associate with it a curve $\bar{\Gamma}$ in flat Minkowski space-time which has the same equations $x^m(s)$ in rectangular Cartesians. The curve $\bar{\Gamma}$ is determined by Γ uniquely to within Lorentz transformations. The principle of equivalence associates a thin tube around Γ with a thin tube around $\bar{\Gamma}$.

Within the framework of general relativity we can now state the principle of equivalence more precisely:

Along any time-like world line, the laws of physics (other than those of gravitation) have the same form in Fermi coordinates as they do in special relativity theory in inertial frames.

This, together with a criterion of simplicity, determines the physical laws of general relativity theory. The criterion of simplicity requires that no unnecessary terms be added which contain gradients of the gravitational field, i.e., second derivatives of the metric tensor, and which vanish in flat Minkowski space-time. This means that no unnecessary terms are added which contain the curvature tensor $R^r{}_{smn}$. We illustrate this with the set (2.12) of Maxwell's equations. We write

$$(3.8) \qquad F^{mn}{}_{;n} = 4\pi J^m,$$

where the semicolon denotes covariant differentiation. This equation reduces to $F^{mn}{}_{,n} = 4\pi J^m$ along Γ in Fermi coordinates. The criterion of simplicity eliminates equations such as $F^{mn}{}_{;n} = 4\pi J^m + R^m{}_{nrs} J^n F^{rs}$.

Table I provides a summary of the tensor calculus.[8] Table II gives most of the basic laws of physics in general relativity theory. One important law which is missing is the relativistic analogue of $\Delta V = 4\pi\rho$, the gravitational field equations. This is left to the next chapter.

TABLE I. Summary of Tensor Calculus

Dual of skew symmetric tensor:

$$\eta_{rsmn} = (-g)^{1/2}\epsilon_{rsmn}, \quad \eta^{rsmn} = -\epsilon_{rsmn}/(-g)^{1/2}$$

$$T^*{}_{rs} = \frac{1}{2}\eta_{rsmn}T^{mn}, \quad T^{**}{}_{rs} = -T_{rs}$$

Christoffel symbol:

$$\left\{ \begin{matrix} r \\ m\ n \end{matrix} \right\} = \left\{ \begin{matrix} r \\ n\ m \end{matrix} \right\} = \frac{1}{2}g^{rs}(g_{ns,m} + g_{ms,n} - g_{mn,s}),$$

$$\left\{ \begin{matrix} r \\ m\ r \end{matrix} \right\} = ((-g)^{1/2})_{,m}/(-g)^{1/2}, \quad [mn,r] = g_{rs}\left\{ \begin{matrix} s \\ m\ n \end{matrix} \right\}$$

Absolute derivative along curve $x^m(u)$:

$$\frac{DT^{r\cdots}_{s\cdots}}{du} = \frac{dT^{r\cdots}_{s\cdots}}{du} + \left\{ \begin{matrix} r \\ a\ m \end{matrix} \right\} T^{a\cdots}_{s\cdots}\frac{dx^m}{du} + \cdots - \left\{ \begin{matrix} a \\ s\ m \end{matrix} \right\} T^{r\cdots}_{a\cdots}\frac{dx^m}{du} - \cdots$$

Covariant derivative:

$$T^{r\cdots}_{s\cdots;m} = T^{r\cdots}_{s\cdots,m} + \left\{ \begin{matrix} r \\ a\ m \end{matrix} \right\} T^{a\cdots}_{s\cdots} + \cdots - \left\{ \begin{matrix} a \\ s\ m \end{matrix} \right\} T^{r\cdots}_{a\cdots} - \cdots$$

$$\left(\frac{DT^{r\cdots}_{s\cdots}}{du} = T^{r\cdots}_{s\cdots;m}\frac{dx^m}{du} \right)$$

Parallel propagation along curve $x^m(u)$:

$$\frac{DT^{\cdots}_{\cdots}}{du} = 0$$

TABLE I (continued)

Geodesic:	
a. affine (preferred) parameter u	$\frac{D}{du}\frac{dx^m}{du} = 0 \quad \left(g_{mn}\frac{dx^m}{du}\frac{dx^n}{du} = \text{const}, \; u = s \text{ for nonnull geodesic}\right)$
b. general parameter v	$\frac{D}{dv}\frac{dx^m}{dv} = \lambda(v)\frac{dx^m}{dv}$
c. variational principle	$\delta\int g_{mn}\frac{dx^m}{du}\frac{dx^n}{du}\,du = 0 \quad \left(\delta\int ds = 0 \text{ for nonnull geodesic}\right)$
Riemann (curvature) tensor:	$T_{s;mn} - T_{s;nm} = R^r{}_{smn}T_r$
	$x^r(u,v)$: $\frac{D^2T^r}{\partial u\partial v} - \frac{D^2T^r}{\partial v\partial u} = R^r{}_{smn}T^s\frac{\partial x^m}{\partial u}\frac{\partial x^n}{\partial v}$
	$R^r{}_{smn} = \left\{\begin{matrix} r \\ s\ n \end{matrix}\right\}_{,m} - \left\{\begin{matrix} r \\ s\ m \end{matrix}\right\}_{,n} + \left\{\begin{matrix} a \\ s\ n \end{matrix}\right\}\left\{\begin{matrix} r \\ a\ m \end{matrix}\right\} - \left\{\begin{matrix} a \\ s\ m \end{matrix}\right\}\left\{\begin{matrix} r \\ a\ n \end{matrix}\right\}$
	$R_{rsmn} = \frac{1}{2}(g_{rn,sm} + g_{sm,rn} - g_{rm,sn} - g_{sn,rm}) + g^{ab}([rn,a][sm,b] - [rm,a][sn,b])$
Algebraic symmetries:	$R_{rsmn} = R_{[rs][mn]} = R_{mnrs}, \quad R_{r[smn]} = 0$
Bianchi identity:	$R_{rs[mn;t]} = 0$
Ricci tensor, curvature invariant:	$R_{mn} = R_{nm} = R^r{}_{mnr}, \quad R = R^m{}_m$
Einstein tensor:	$G_{mn} = G_{nm} = R_{mn} - \frac{1}{2}g_{mn}R, \quad R_{mn} = G_{mn} - \frac{1}{2}g_{mn}G$
	$G^m{}_{n;m} = 0$

TABLE I (continued)

Weyl tensor (conformal curvature tensor):	$C^r{}_{smn} = R^r{}_{smn} + g_{s[n}R^r{}_{m]} + R_{s[n}\delta^r{}_{m]} - \frac{1}{3}g_{s[n}\delta^r{}_{m]}R$
	$C_{rsmn} = C_{[rs][mn]} = C_{mnrs}, \quad C_{r[smn]} = 0, \quad C^r{}_{smr} = 0$
	$(\bar{g}_{mn} = e^{2\sigma(x)}g_{mn}: \ \bar{C}^r{}_{smn} = C^r{}_{smn})$

TABLE II. Physical Laws in General Relativity Theory

Free particle (free fall):	$\frac{D}{ds}\frac{dx^m}{ds} = \frac{d^2x^m}{ds^2} + \left\{ \begin{matrix} m \\ r \ s \end{matrix} \right\} \frac{dx^r}{ds}\frac{dx^s}{ds} = 0, \quad ds^2 > 0.$
Light pulse (photon):	$\frac{D}{du}\frac{dx^m}{du} = 0, \quad ds^2 = 0.$
World momentum:	
a. material particle	$p^m = mu^m, \quad u^m = dx^m/ds, \quad u^m u_m = 1, \quad p^m p_m = m^2$
b. photon	$p^m = hk^m, \quad k^m = \nu(n^\mu, 1), \quad n^\mu n^\mu = 1, \quad k^m k_m = 0$
Collision laws:	$\Sigma p^m = \bar{\Sigma}\bar{p}^m$
World force:	$\frac{Dp^m}{ds} = F^m, \quad F^m u_m = 0$
Electromagnetism:	$F_{mn} = -F_{nm}, \quad F_{[mn,r]} = 0 \Leftrightarrow F^{*mn}{}_{;n} = 0$
a. Maxwell's Equations	$\Leftrightarrow ((-g)^{1/2}F^{*mn})_{,n} = 0,$
	$F^{mn}{}_{;n} = 4\pi J^m \Leftrightarrow ((-g)^{1/2}F^{mn})_{,n} = 4\pi(-g)^{1/2}J^m$
b. Electromagnetic potentials	$F_{mn} = 2A_{[m,n]},$
	$A^{m\,n}{}_{;n} - A^{n}{}_{;n}{}^{m} + R^{nm}A_n = 4\pi J^m$
c. Charge of isolated system	$J^m{}_{;m} = 0 \Leftrightarrow ((-g)^{1/2}J^m)_{,m} = 0$
	$e = \int_{\tau(3)} (-g)^{1/2}J^m d\tau_{(3)m}$ (cf. Figure 10)

TABLE II (continued)

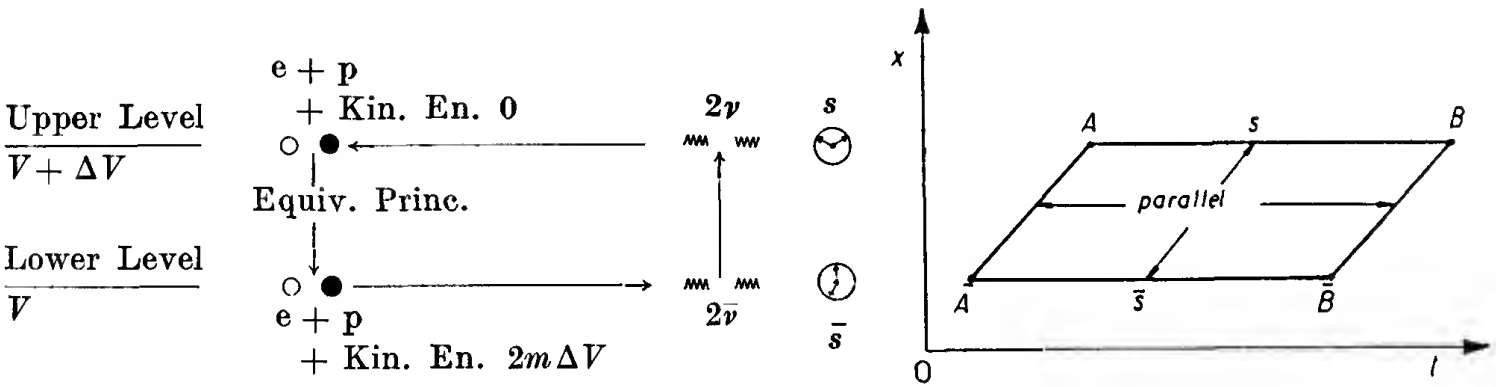
d. Lorentz equations of motion	$F^m = eu^n F_n{}^m, \quad F^m u_m = 0$ $m\frac{D}{ds}\frac{dx^m}{ds} = e\frac{dx^n}{ds}F_n{}^m \Leftrightarrow \delta\int (mds + eA_m dx^m) = 0$
e. Maxwell stress energy tensor	$E^{mn} = \frac{1}{4\pi}\left[-F^m{}_a F^{na} + \frac{1}{4} g^{mn} F^{ab} F_{ab}\right],$ $E^{mn}{}_{;n} = -J^n F_n{}^m$
Material stress energy tensor:	$P^{mn} = P^{nm}$
a. Perfect fluid	$P^{mn} = (\tilde{\rho} + \tilde{p})\, u^m u^n - \tilde{p} g^{mn}, \quad u^m u_m = 1$
b. Dust	$P^{mn} = \tilde{\rho}\, u^m u^n$
Total stress energy tensor:	$T^{mn} = T^{nm}, \quad T^{mn}{}_{;n} = 0$
a. Matter and electromagnetic field	$T^{mn} = P^{mn} + E^{mn}$

One important remark must be made about the first line of Table II. By a free particle in general relativity theory we now understand what in classical physics would be called a particle under no forces other than gravitation. Gravitation is no longer regarded as a field which exerts forces; it is the geometry of space-time.

The Principle of Equivalence; Energy Argument. The theory of general relativity is exactly fifty years old. For a theory of such venerable age, the experimental and observational confirmation is remarkably meager. This is because the gravitational field is weak, and most gravitational systems on the astronomical scale are non-relativistic.

By far the most accurate verification of the principle of equivalence, and of its immediate consequences, is the gravitational red shift.

[8] J. L. Synge and A. Schild, *Tensor calculus,* University of Toronto Press, Toronto, 1964. We assume that the reader knows the contents of this book.



(a)

$$2h\nu = 2m$$

$$2h\bar{\nu} = 2m + 2m\,\Delta V$$

$$\therefore \frac{\bar{\nu} - \nu}{\nu} = \Delta V$$

(b)

$$\nu = n/s$$

$$\bar{\nu} = n/\bar{s}$$

$$\therefore \frac{s - \bar{s}}{\bar{s}} = \Delta V$$

(c)

$$\bar{s} \neq s$$

$\therefore$ space-time is curved

FIGURE 14

This last effect has been directly measured by Pound and Rebka[9] and Pound and Snider[10] with an accuracy of about 1%, and indirectly by the Dicke-Eötvös experiments which weigh electrostatic fields and virtual photons of nuclei.[11]

We shall now give a heuristic argument which goes a long way to show that the principle of equivalence leads to a curved space-time theory of gravitation with the features discussed in this chapter.[12]

Consider a spherically symmetric gravitating body of mass M at rest in an otherwise empty universe. Near its surface it will produce a static gravitational field which, in Newtonian approximation, has the potential

$$V = -\frac{M}{r}. \tag{3.9}$$

The first part of the argument is illustrated by Figure 14a. An electron e and a positron p, each of mass m, are at rest at a level in the static gravitational field where the potential is $V + \Delta V$. Both particles are permitted to fall to a lower level where the potential is V. The equivalence principle states that both particles accelerate down together and each acquires a kinetic energy $m\Delta V$. The moving electron and positron at the lower level V are permitted to annihilate there and produce two photons of frequency $\bar{\nu}$. By means of elastic scattering from heavy mirrors the two photons are directed upwards to the original level $V + \Delta V$, where they have frequency ν. Finally the cycle is completed by permitting the two photons ν to create an electron-positron pair at level $V + \Delta V$. The final pair can be produced by the two photons and it must again be at rest, as was the initial pair. Otherwise it would be possible to run the cycle, or its reverse, in such a way that energy would be produced from nothing. This would constitute a perpetuum mobile, and we accept the empirical fact that this is impossible. For the reactions at two levels of our cycle we have the equations

$$2h\nu = 2m, \qquad 2h\bar{\nu} = 2m + 2m\Delta V, \tag{3.10}$$

---

[9] R. V. Pound and G. A. Rebka, Jr., Phys. Rev. Letters 4 (1960), 337.
[10] R. V. Pound and J. L. Snider, Phys. Rev. Letters 13 (1964), 539.
[11] L. I. Schiff, Proc. Nat. Acad. Sci. U.S.A. 46 (1960), 871.
[12] A. Schild, Proc. Internat. School of Physics "Enrico Fermi," Academic Press, New York, 1963; Course 20, pp. 69-115.

where h is Planck's constant. It immediately follows that

$$\frac{\bar{\nu} - \nu}{\nu} = \Delta V. \tag{3.11}$$

This is the formula for the *gravitational red shift.* It states that light moving up in a gravitational field must change in frequency from $\bar{\nu}$ to ν and thus become redder.

The second part of the argument is illustrated by Figure 14b. We now consider the light moving up from level V to $V + \Delta V$, not from the point of view of a quantum phenomenon, but from the dual point of view of a wave phenomenon. Let us think of a radio station at level V sending out a continuous monochromatic radio signal of frequency $\bar{\nu}$. A receiver, resting at level $V + \Delta V$, receives the signal at the lower frequency ν. We now have an apparent contradiction: Since the whole setup is stationary, and since wave crests cannot originate or disappear between sender and receiver, it seems that a change in frequency is impossible. The way out of this difficulty is to assume that time flows at different rates at different levels in a gravitational field. If a clock resting at level V measures a time interval $\bar{s}$ for n oscillations of the radio wave and if a clock resting at level $V + \Delta V$ measures a time interval s for n oscillations, then $\nu = n/s$ and $\bar{\nu} = n/\bar{s}$. Substituting this into the gravitational red shift formula (3.11) we obtain

$$\frac{s - \bar{s}}{\bar{s}} = \Delta V. \tag{3.12}$$

This is the formula for the *gravitational time dilatation.* It states that a clock resting at a higher level in a static gravitational field goes faster than an identical clock resting at a lower level.

The third part of the argument is illustrated by Figure 14c. We begin by assuming that special relativity is valid and that space-time is Minkowskian with line element $ds^2 = -dx^2 - dy^2 - dz^2 + dt^2$, where ds is the element of proper time measured by a physical clock along its world line. Our heavy, gravitating body M is to be initially at rest in the inertial frame x, y, z, t. Then, independently of any assumptions about the gravitational field, it follows from the spherical symmetry of M that it will be permanently at rest, and that the world line of any particle resting near the surface of M will be a straight line parallel to the t-axis. Figure

14c shows a space-time diagram of the two clocks at levels V and $V + \Delta V$ (to preserve continuity with Figure 14a and 14b, the space axis has been drawn vertically and the time axis horizontally). AB is the world line of the lower clock and $\bar{A}\bar{B}$ that of the higher clock; the two world lines are parallel. $A\bar{A}$ is the world line of a light signal used to compare the readings of the two clocks at the beginning of the experiment, $B\bar{B}$ that of a light signal used to compare the clocks at the end of the experiment. Since the gravitational field is static, it follows that the propagation properties of light from level V to $V + \Delta V$ are independent of the time when the light was emitted; they are the same for the light signals $A\bar{A}$ and $B\bar{B}$. Therefore the lines $A\bar{A}$ and $B\bar{B}$ are parallel. In flat Minkowski space-time it now follows that $AB\bar{B}\bar{A}$ is a parallelogram and that the two sides $\bar{A}\bar{B} = \bar{s}$ and $AB = s$ are equal. However, the gravitational time dilatation effect, Equation (3.12), states precisely that $\bar{s}$ and s are not equal.

We have now arrived at a contradiction. We conclude that the equivalence principle, the gravitational red shift or the gravitational time dilatation are inconsistent with special relativity theory.

One can go further. In geometry the absence of parallelism and of parallelograms with the usual properties is characteristic of a curved space. Therefore the results of this section provide a strong heuristic argument which leads from the principle of equivalence to the conclusion that space-time is a curved Riemannian manifold whose line element ds is the element of proper time measured by a physical clock along its world line.

In the argument of this section, particularly in the first part illustrated by Figure 14a, a number of assumptions were made implicitly. It was assumed, for example, that no energies depending on the gravitational potential V enter the Equations (3.10) and that Planck's constant h is in fact a constant and independent of the gravitational potential V. Since the gravitational potential V is physically determined only to within an arbitrary additive constant, whereas the frequency ν of a photon or Planck's constant h do not permit the addition of such an arbitrary constant, these assumptions are reasonable. However, we shall show in the next section that these assumptions are extraneous and unnecessary. We shall give a more direct and more precise derivation of the gravitational time dilatation from the principle of equivalence.

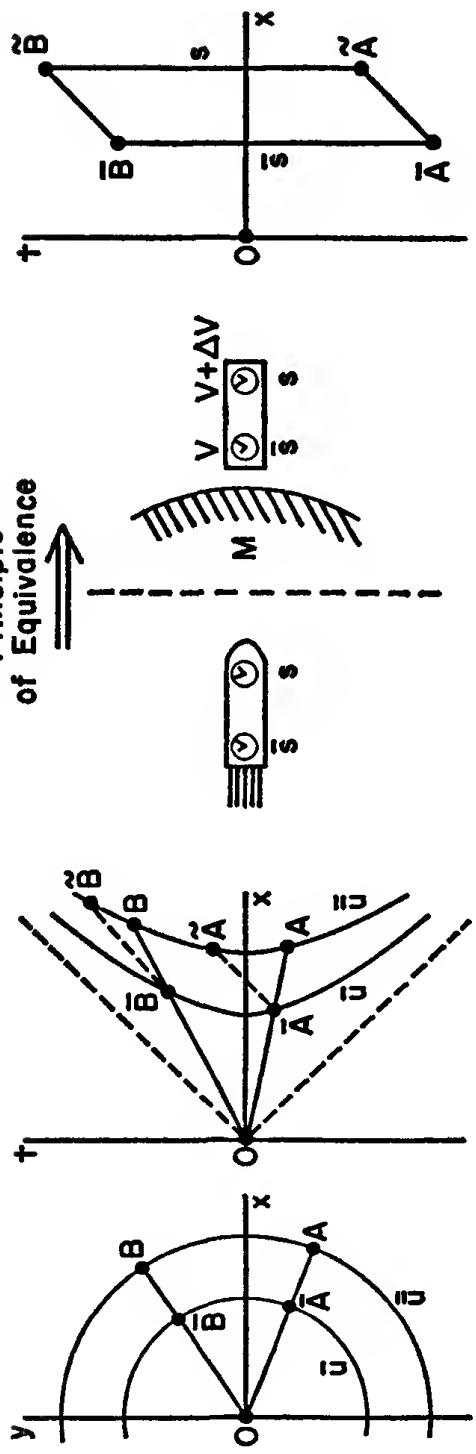

$x^2 + y^2 = u^2$
Euclidean Space
(a)
$x^2 - t^2 = u^2$
Minkowski Space–Time
(b)
Accelerated Elevator in Space
(c)
Gravitational Field in Space
(d)
$\bar{s} \neq s$
∴ Space–Time is Curved
(e)
Principle of Equivalence
M
V
V+ΔV
s
x
t
y
O

FIGURE 15

The Principle of Equivalence; The Elevator Argument. In Minkowski space-time, free of gravitation, consider an elevator accelerated along the positive x-axis of an inertial frame. A simple relativistic model of such an elevator is a system of particles, each corresponding to a constant value of the parameter u, whose world lines are given by the concentric circles (which are hyperbolae in Minkowski space):

$$x^2 - t^2 = u^2, \qquad \bar{u} \leqq u \leqq \bar{\bar{u}}. \tag{3.13}$$

This is shown in Figure 15b. The following facts can be deduced directly from the geometry of Minkowski space. They can also be easily seen by analogy with a system of concentric circles in Euclidean space

$$x^2 + y^2 = u^2, \qquad \bar{u} \leqq u \leqq \bar{\bar{u}}, \tag{3.14}$$

which are shown in Figure 15a.

Any radial line, such as $O\bar{A}A$, cuts all the circles of the system orthogonally. Physically, this means that the radial lines are lines of simultaneity within the elevator.

The intercepts between any two circles of the system are the same for all radial lines, e.g., $\bar{A}A = \bar{B}B$. Physically, this means that the elevator is rigid.

Each circle has a constant curvature g which is the reciprocal of its radius u, $g = 1/u$. Physically, this means that each particle of the elevator has an intrinsic acceleration g which remains constant throughout its history. The intrinsic acceleration is the ordinary Newtonian acceleration measured by an inertial observer instantaneously at rest relative to the particle. The intrinsic acceleration varies from particle to particle, but we can ignore this variation by making the elevator small: $\bar{\bar{u}} - \bar{u} \ll \bar{u}$. In this approximation the rigid elevator has constant uniform acceleration

$$g = 1/\vec{u}. \tag{3.15}$$

We shall call the small quantity

$$\Delta V = g(\bar{A}A) = g(\bar{\bar{u}} - \bar{u}) \tag{3.16}$$

the difference in acceleration potential between the top (A) and the bottom ($\bar{A}$) of the elevator.

The arc length along a circle intercepted between two fixed

radial lines is proportional to the radius of the circle. Thus if $\bar{s}$ is the arc $\overline{AB}$ and s the arc AB, $s/\bar{s} = \bar{\bar{u}}/\bar{u}$. In the approximation of a small elevator, this gives

$$\frac{s - \bar{s}}{\bar{s}} = \Delta V. \tag{3.17}$$

Physically, this means that clocks at different levels of the elevator go at different rates, the comparison of the clocks being by means of simultaneity within the elevator. We shall call this effect the *acceleration time dilatation.*

The system of circles is invariant under rotations about 0. Physically this means that the system is stationary, and that therefore the acceleration time dilatation formula (3.17) remains valid whatever the method used for comparing the two clocks, provided only that the same intrinsic method is used at the beginning and at the end of the experimental run. Let us, for example, consider the comparison of the two clocks by means of light signals $\bar{A}\,\widetilde{A}$ and $\bar{B}\,\widetilde{B}$ (null lines in Figure 15b) sent from the lower to the upper clock. A suitable rotation about 0 (a Lorentz transformation) will take $\bar{A}A\,\widetilde{A}$ into $\bar{B}B\,\widetilde{B}$, and therefore $\operatorname{arc} A\,\widetilde{A} = \operatorname{arc} B\,\widetilde{B}$. It follows that

$$\operatorname{arc} \widetilde{A}\,\widetilde{B} = \operatorname{arc} AB = s, \tag{3.18}$$

and that Equation (3.17) holds when the clocks are compared by light signals moving up in the elevator.

It is now easy to see that (3.17) implies an *acceleration red shift* formula

$$\frac{\bar{\nu} - \nu}{\nu} = \Delta V, \tag{3.19}$$

for light moving up in the elevator.

So far everything stated above is a direct consequence of special relativity theory and since special relativity is amply verified, we accept it completely.

It is at this point that we use the principle of equivalence. In the elevator, any body which is released moves relative to the

elevator with an acceleration g along the negative x-axis. The equivalence principle states that all physical phenomena within the elevator (Figure 15c) are the same as in a small region in a static gravitational field (Figure 15d) where the acceleration of a freely falling body is g.

It is the principle of equivalence which permits us to conclude from the acceleration time dilatation that two clocks resting at different levels in a static gravitational field (Figure 15d) go at different rates, and that this gravitational time dilatation is given by Equation (3.17). Similarly we conclude that there is a gravitational red shift given by Equation (3.19).

We have now gained the same result which we reached by the second part of the argument of the previous section. We can continue exactly as before and find that the space-time in the neighborhood of a heavy, spherically symmetric body must be curved.

There is a second, and equivalent, way of seeing that space-time must be curved. If we consider the history of a single small region R near a spherically symmetric heavy mass M at rest (Figure 16a), then no physical experiment within R can distinguish between the gravitational field in R and the case of an accelerated elevator E (Figure 16b) in flat space-time. The situation changes as soon as we consider several small regions $R, R', R'', \cdots$ (Figure 16a) which fill a thin hemispherical shell S. The time dilatation within each of the regions would tell us that, if space-time were flat, the regions would be accelerating radially as do the elevators $E, E', E'', \cdots$ in Figure 16b. Now there is a sharp distinction between the gravitational case (a) which is stationary, and the acceleration case (b) which is not. In Figure 16a the distance between neighboring regions R and R' does not change, e.g., the time for a light signal to travel from R to R' and back is independent of the time of emission. In Figure 16b the distance between neighboring elevators E and E' changes, e.g., the time for a light signal to travel from E to E' and back varies with the time of emission. It is therefore impossible to represent the physical observations within the shell S, and throughout its history, in a flat space-time. The space-time surrounding a heavy, spherically symmetric body must be curved.

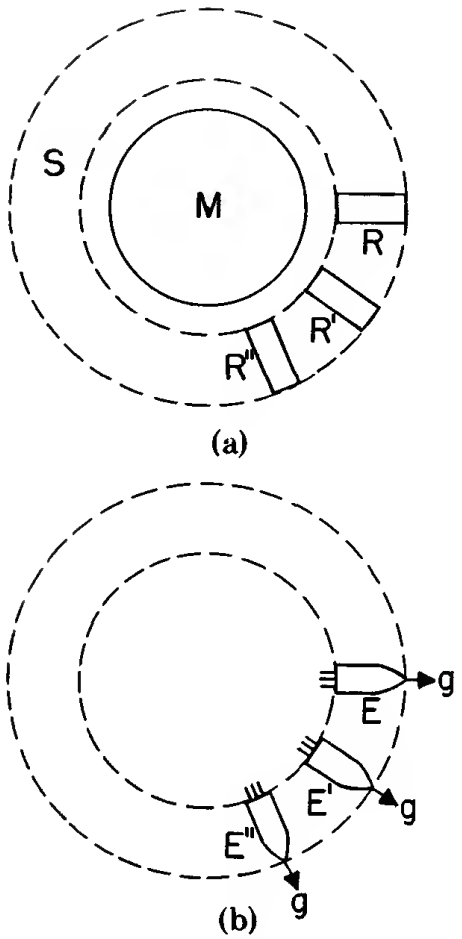


FIGURE 16

The arguments of the last two sections show that the principle of equivalence is incompatible with the theory of special relativity. Thus, unless artificial and ad hoc assumptions are admitted, gravitation cannot be described by a standard field theory in Minkowski space.

Exercise. Consider two particles Γ and $\bar{\Gamma}$ which are rigidly connected and have neighboring motions. This means that the orthogonal connecting vector δx^m (Figure 17) between their world lines has constant length. Show that Γ is a geodesic if and only if there is zero time dilatation between Γ and all neighboring particles $\bar{\Gamma}$ rigidly connected to Γ, the method of comparison of times being either by intrinsic simultaneity ($A\bar{A}$ and $B\bar{B}$) or by light signals ($\bar{A}\tilde{A}$ and $\bar{B}\tilde{B}$). Apply the equivalence principle to show that a particle falling freely in a gravitational field has a time-like geodesic as world line.

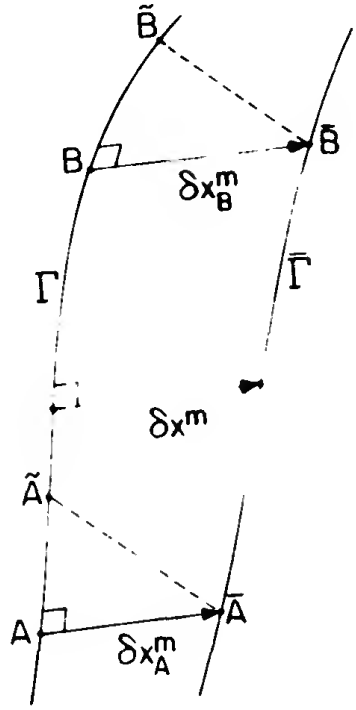


FIGURE 17

4. **The gravitational field equations.** In Newtonian gravitational theory the basic equations are the equations of motion

$$\frac{d^2x^\mu}{dt^2} = -V_{,\mu}, \tag{4.1}$$

and the field equation

$$V_{,\mu\mu} = 4\pi\rho. \tag{4.2}$$

In Einstein's theory of general relativity the equations of motion of a test particle are those of a geodesic (Table II, line 1). The Christoffel symbol term contains first derivatives of g_{mn} and is analogous to the gradient term $V_{,\mu}$ in Equation (4.1).

The source term ρ of Newtonian theory is just one component of the second order tensor T_{mn} in relativity theory. We may therefore expect that the whole stress energy tensor will appear as the source term in the relativistic field equations for gravitation, and that these equations will be second order tensor equations. The *gravitational field equations* of general relativity theory are

$$G_{mn} = \alpha T_{mn}. \tag{4.3}$$

In *vacuum*, when the source term vanishes, $T_{mn} = 0$, the field equa-

tions are

$$G_{mn} = 0 \Leftrightarrow R_{mn} = 0. \tag{4.4}$$

These equations have the following properties:

1. Like the Newtonian field equations, they are second order partial differential equations in the gravitational potentials g_{mn}. They are, however, nonlinear.

2. Since the Einstein tensor has zero divergence, $G^{m}{}_{n,m} = 0$, the differential conservation laws

$$T^{mn}{}_{;m} = 0 \tag{4.5}$$

are consequences of the field equations.

3. In the nonrelativistic approximation, the equations reduce to the Newtonian field equations. Thus general relativity theory reproduces all the results of Newtonian gravitation in first approximation. This will be shown in this chapter. The argument will determine the value of the coupling constant α.

4. Flat Minkowski space, where the full Riemann tensor vanishes,

$$R^{m}{}_{nrs} = 0, \tag{4.6}$$

is a particular solution of the vacuum equations (4.4).

5. The field equations can be obtained from a variational principle. This will be shown in Chapter 5.

6. General relativity theory has an important property of consistency. The equations of motion are not independent of the field equations, but are to a large extent determined by them. This will be discussed in Chapter 6.

7. The gravitational field, like other fields of zero proper mass, propagates with the velocity of light. This will be discussed in Chapter 7.

The field equations (4.3) are unique in having all these properties.

Newtonian Approximation. We assume that the gravitational field is weak and quasi-static, and that particles are slow.

The weakness of the gravitational field means that space-time is almost flat. Coordinates exist such that

$$g_{mn} = \eta_{mn} + h_{mn}, \tag{4.7}$$

where h_{mn} and its derivatives are small. Products of two of these small quantities will be neglected.

Quasi-static means that the time derivatives $g_{mn,4} = h_{mn,4}$ are small compared to $h_{mn,\rho}$ and will be neglected. This corresponds to the fact that in a nonrelativistic system the change in the gravitational potential over some distance is large compared to the change of the potential at one point in the short time it takes light to traverse this distance.

Slow means that speeds of particles are small compared to that of light.

We first study the geodesic equations of motion. Only three of them are independent, since

$$\frac{dx_m}{ds}\frac{D}{ds}\frac{dx^m}{ds} \equiv 0, \tag{4.8}$$

and we may take these to be

$$\frac{d^2x^\mu}{ds^2} + \left\{ \begin{matrix} \mu \\ r\ s \end{matrix} \right\} \frac{dx^r}{ds}\frac{dx^s}{ds} = 0. \tag{4.9}$$

In these equations d^2x^μ/ds^2 and

$$\left\{ \begin{matrix} \mu \\ r\ s \end{matrix} \right\}$$

are already small. We may therefore write in the Newtonian approximation

$$\frac{d^2x^\mu}{dt^2} = -\left\{ \begin{matrix} \mu \\ 4\,4 \end{matrix} \right\} = \frac{1}{2}\eta^{\mu\nu}g_{44,\nu} = -\frac{1}{2}g_{44,\mu}. \tag{4.10}$$

This agrees with Equations (4.1) if we identify

$$g_{44} = 1 + 2V. \tag{4.11}$$

We next study the field equations and consider the trial solution

$$h_{mn} = 2\delta_{mn}V, \qquad g_{mn} = \eta_{mn} + 2\delta_{mn}V. \tag{4.12}$$

A straightforward calculation shows that in the Newtonian approximation

$$R_{mn} = -\delta_{mn}V_{,\rho\rho}. \tag{4.13}$$

The field equations (4.3) now become

$$G_{44} = \alpha T_{44} = \alpha\rho = -2V_{,\rho\rho}, \tag{4.14a}$$

$$G_{\mu 4} = \alpha T_{\mu 4} = 0, \qquad G_{\mu\nu} = \alpha T_{\mu\nu} = 0. \tag{4.14b}$$

In a nonrelativistic system, density, velocity and stress are of order of magnitude one in cgs units. A simple dimensional analysis then shows that in our units ($c = G = 1$) all other components of T_{mn} are small compared to T_{44}. This explains the Equations (4.14b) in the Newtonian approximation.

Equation (4.14a) agrees in form with the Newtonian field equation (4.2), and determines the coupling constant $\alpha = -8\pi$. Thus the relativistic field equations are

$$G_{mn} = -8\pi T_{mn}. \tag{4.15}$$

The analysis of this section shows that every nonrelativistic motion of a Newtonian gravitating system can be reproduced as a first approximation of the relativistic laws.

Equations of Geodesic Deviation. The acceleration of one freely falling particle has no intrinsic significance in general relativity theory. The particle is at rest for an observer moving with it. The relative acceleration of two neighboring particles, on the other hand, describes intrinsic properties of the gravitational field, the field gradient physically, components of the Riemann tensor geometrically. We shall see that this relative acceleration satisfies equations very similar to those of Newtonian theory.

Consider two neighboring time-like geodesics (Figure 18a), and embed them in a one parameter family of geodesics $x^m(s, v)$; v labels the different geodesics of the family, s is the arc length along each geodesic. Then

$$\frac{Du^m}{\partial s} = 0, \qquad u^m = \frac{\partial x^m}{\partial s}. \tag{4.16}$$

The infinitesimal connecting vector η^m is given by

$$\eta^m = \frac{\partial x^m}{\partial v}\, dv. \tag{4.17}$$

Hence

$$\frac{D\eta^m}{\partial s} = \left[\frac{\partial^2 x^m}{\partial s \partial v} + \begin{Bmatrix} m \\ r\ s \end{Bmatrix} \frac{\partial x^r}{\partial v}\frac{\partial x^s}{\partial s}\right] dv = \frac{Du^m}{\partial v}\, dv. \tag{4.18}$$

We have

$$\frac{D}{\partial s}(\eta^m u_m) = u_m \frac{D\eta^m}{\partial s} = u_m \frac{Du^m}{\partial v}\, dv = \frac{1}{2}\frac{D}{\partial v}(u_m u^m)\, dv = 0. \tag{4.19}$$

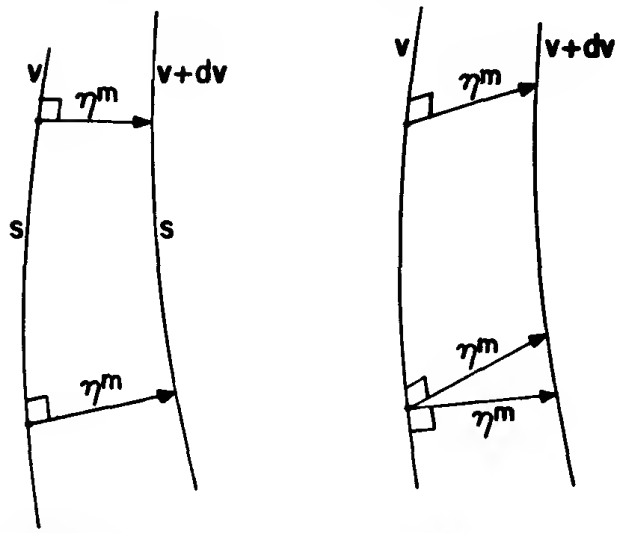


FIGURE 18

We now assume that η^m is initially orthogonal to one of the geodesics:

$$\eta^m u_m = 0. \tag{4.20}$$

It then remains orthogonal all along the geodesic.

Differentiating Equation (4.18), we obtain

$$\frac{D^2\eta^m}{\partial s^2} = \frac{D}{\partial s}\frac{Du^m}{\partial v}dv = \frac{D}{\partial v}\frac{Du^m}{\partial s}dv + R^m{}_{nrs}u^n u^r \eta^s.$$

Using Equation (4.16), we obtain the equations of geodesic deviation

$$\frac{D^2\eta^m}{ds^2} = R^m{}_{nrs}u^n u^r \eta^s. \tag{4.21}$$

We have reverted to the notation of ordinary (rather than partial) absolute derivatives in order to stress that we can now forget about the one parameter family of curves and that these equations apply to two neighboring geodesics.

Exercise. Show that, if two neighboring null geodesics have one orthogonal connecting vector, then all connecting vectors are

orthogonal (Figure 18b). Show also that the Equations (4.21) of geodesic deviation apply to neighboring null geodesics if the arc length s is replaced by an affine parameter u.

Consider a cloud of test particles moving in a gravitational field. Let x^μ and $x^\mu + \eta^\mu$ be the space coordinates of two neighboring particles. The Newtonian equations of motion (4.1) give the relative acceleration

$$\frac{d^2\eta^\mu}{dt^2} = K_{\mu\nu}\eta^\nu, \tag{4.22}$$

where $K_{\mu\nu}\eta^\nu = -(V_{,\mu})_{x+\eta} + (V_{,\mu})_x$, or

$$K_{\mu\nu} = -V_{,\mu\nu}. \tag{4.23}$$

The field equation (4.2) can now be written

$$K_{\mu\mu} = -4\pi\rho. \tag{4.24}$$

In relativity theory, the cloud of particles have as world lines a congruence of time-like geodesics. Introduce geodesic Fermi coordinates based on one of these geodesics. Then the equations of geodesic deviation for neighboring particles become

$$\frac{d^2\eta^\mu}{dt^2} = R^\mu{}_{44\nu}\eta^\nu. \tag{4.25}$$

This has the same form as the Newtonian Equation (4.22) with

$$K_{\mu\nu} = R^\mu{}_{44\nu}. \tag{4.26}$$

From the relativistic field Equations (4.15), we obtain

$$K_{\mu\mu} = R_{44} = -8\pi\left(T_{44} - \frac{1}{2}T\right). \tag{4.27}$$

This is similar to Equation (4.24). In fact, in the nonrelativistic approximation $T = T_{44} = \rho$, and the right-hand side of Equation (4.27) becomes $-4\pi\rho$, in complete agreement with the Newtonian result.

The Cosmological Term. On the cosmological scale, field equations more general than (4.15) are sometimes used:

$$G_{mn} + \Lambda g_{mn} = -8\pi T_{mn}. \tag{4.28}$$

The left-hand side again has zero divergence, so that the conserva-

tion equations $T^{mn}{}_{;m} = 0$ follow. The *cosmological constant* Λ must be very small. Otherwise Newtonian gravitational theory cannot be obtained in first approximation. Except for the following chapter, we shall assume that the cosmological constant is zero.

5. **Variational principle.** In this chapter we shall study a pure field theory, with the gravitational field described by the variables g^{mn} and the sources of the gravitational field described by other field variables y_μ, Greek suffixes ranging and summing over 1, $2, \cdots, N$. We shall restrict ourselves to the case where the y_μ are a collection of tensor components. The general case, where spinors are included, is more complicated and has been discussed by Rosenfeld.[13]

The field equations are obtained from a variational principle with an action

$$I = \int_{\tau(4)} (-g)^{1/2} L d\tau_{(4)}, \qquad L = L_{(G)} + L_{(F)}. \tag{5.1}$$

The Lagrangians L, $L_{(G)}$ and $L_{(F)}$ are chosen to be invariants, since $(-g)^{1/2} d\tau_{(4)}$ is also an invariant. The gravitational Lagrangian is the simplest invariant available

$$L_{(G)} = R - 2\Lambda, \tag{5.2}$$

Λ being a constant, the cosmological constant. The Lagrangian $L_{(F)}$ of the other fields is a function of the variables g^{mn}, y_μ and their first derivatives with respect to the coordinates. A concrete example is the electromagnetic field: The field variables y_μ are the electromagnetic potentials A_m, and the Lagrangian is

$$L_{(F)} = -F_{mn}F^{mn} = -g^{mr}g^{ns}F_{mn}F_{rs}, \tag{5.3}$$

where F_{mn} is understood to stand for

$$F_{mn} = A_{m,n} - A_{n,m}, \tag{5.4}$$

so that one set of Maxwell's equations is automatically satisfied.

No invariant, other than a constant, can be formed from the metric tensor and its first derivatives. Therefore the gravitational Lagrangian must be a second order expression. In general, the vari-

[13] L. Rosenfeld, Mém. Acad. Roy. Belg. No. 6, 18 (1940).

ational principle would then give fourth order field equations. This would be the case if, for example, $L_{(G)}$ contained a term $R^{mnrs}R_{mnrs}$. The Lagrangian chosen, a linear combination of R and a constant, has the unique property of giving second order field equations. This is so because R is linear in the second derivatives of the metric tensor.

We shall now calculate the variation of the action for variations δg^{mn}, δy_μ of the field variables. Throughout, the variations δg^{mn}, $\delta g^{mn}{}_{,r}$, δy_μ are assumed to vanish on the boundary $\tau_{(3)}$ of the four-dimensional region $\tau_{(4)}$.

The work is simplified by noting (Palatini) that Christoffel symbols transform according to

$$\left\{ \begin{matrix} r \\ m\ n \end{matrix} \right\}' = \frac{\partial x'^r}{\partial x^s} \frac{\partial x^a}{\partial x'^m} \frac{\partial x^b}{\partial x'^n} \left\{ \begin{matrix} s \\ a\ b \end{matrix} \right\} + \frac{\partial x'^r}{\partial x^s} \frac{\partial^2 x^s}{\partial x'^m \partial x'^n}, \tag{5.5}$$

so that Christoffel symbols are not tensors, but the difference of two Christoffel symbols (calculated from two metrics) is a tensor. In particular the variation

$$\delta \left\{ \begin{matrix} r \\ m\ n \end{matrix} \right\}$$

is a tensor. At the origin of geodesic coordinates, where

$$\left\{ \begin{matrix} r \\ m\ n \end{matrix} \right\} = 0$$

and $g_{mn,r} = 0$, we have

$$(-g)^{1/2} g^{mn} \delta R_{mn} = \left[(-g)^{1/2} \left(g^{mn} \delta \left\{ \begin{matrix} a \\ m\ a \end{matrix} \right\} - g^{ab} \delta \left\{ \begin{matrix} n \\ a\ b \end{matrix} \right\} \right) \right]_{,n}. \tag{5.6}$$

This is a tensor equation, since for a contravariant vector $(-g)^{1/2} T^n{}_{,n} = ((-g)^{1/2} T^n)_{,n}$. It is therefore valid in any coordinate system.

We have

$$\begin{aligned} \delta I_{(G)} &= \int_{\tau(4)} \delta[(-g)^{1/2}(g^{mn} R_{mn} - 2\Lambda)]\, d\tau_{(4)} \\ &= \int_{\tau(4)} [R_{mn} \delta((-g)^{1/2} g^{mn}) - 2\Lambda \delta(-g)^{1/2}]\, d\tau_{(4)} \\ &\quad + \int_{\tau(4)} (-g)^{1/2} g^{mn} \delta R_{mn}\, d\tau_{(4)}. \end{aligned} \tag{5.7}$$

By Equation (5.6) and Green's theorem, the last integral becomes an integral over the boundary $\tau_{(3)}$ and therefore vanishes, since the variations are zero on $\tau_{(3)}$. The other integral is easily evaluated by using $\delta(-g)^{1/2} = \frac{1}{2}(-g)^{1/2}g^{mn}\delta g_{mn} = -\frac{1}{2}(-g)^{1/2}g_{mn}\delta g^{mn}$. The result gives the variation of the gravitational action:

$$\delta I_{(G)} = \int_{\tau(4)} (-g)^{1/2}(G_{mn} + \Lambda g_{mn})\, \delta g^{mn} d\tau_{(4)}. \tag{5.8}$$

The variation of the field action has the form

$$\delta I_{(F)} = \int_{\tau(4)} (-g)^{1/2}[8\pi T_{mn}\delta g^{mn} + L_{(F)}{}^{\mu}\delta y_{\mu}]d\tau_{(4)}, \tag{5.9}$$

where

$$8\pi(-g)^{1/2}T_{mn} = \frac{\partial(-g)^{1/2}L_{(F)}}{\partial g^{mn}} - \left(\frac{\partial(-g)^{1/2}L_{(F)}}{\partial g^{mn}{}_{,r}}\right)_{,r}, \tag{5.10}$$

$$(-g)^{1/2}L_{(F)}{}^{\mu} = \frac{\partial(-g)^{1/2}L_{(F)}}{\partial y_{\mu}} - \left(\frac{\partial(-g)^{1/2}L_{(F)}}{\partial y_{\mu,r}}\right)_{,r}. \tag{5.11}$$

If the field action is purely electromagnetic, and given by Equation (5.3), then its variation is

$$\delta I_{(F)} = \int_{\tau(4)} [8\pi(-g)^{1/2}E_{mn}\delta g^{mn} + 4((-g)^{1/2}F^{mn})_{,n}\delta A_m]d\tau_{(4)}, \tag{5.12}$$

where E_{mn} is the Maxwell stress energy tensor given in Table II.

The Field Equations. The field equations are obtained from the variational principle or action principle

$$\delta I = 0$$

for variations δg^{mn}, δy_{μ} which are arbitrary in $\tau_{(4)}$, but which vanish (together with $\delta g^{mn}{}_{,r}$) on the boundary $\tau_{(3)}$. This immediately gives the gravitational field equations

$$G_{mn} + \Lambda g_{mn} + 8\pi T_{mn} = 0 \tag{5.13}$$

of general relativity theory, and the equations for the other fields

$$L_{(F)}{}^{\mu} = 0. \tag{5.14}$$

Thus the expression given by Equation (5.10) determines the total stress energy tensor T_{mn} in terms of the field variables.

In the purely electromagnetic case, $T_{mn} = E_{mn}$ and the electromagnetic field equations are Maxwell's equations

$$((-g)^{1/2}F^{mn})_{,n} = 0. \tag{5.15}$$

Exercise. In an *affine space,* a set of quantities $\Gamma_{m\,n}^{\ r} = \Gamma_{n\,m}^{\ r}$ are defined which are called *coefficients of connection* and which transform like the Christoffel symbols in Equation (5.5). In such a space covariant differentiation and the curvature tensors $R^{m}{}_{nrs}$, R_{mn} can be defined exactly as in a Riemannian space, $\Gamma_{m\,n}^{\ r}$ replacing

$$\begin{Bmatrix} r \\ m\ n \end{Bmatrix}.$$

Consider the action in Equation (5.1) where $L_{(G)} = g^{mn}R_{mn} - 2\Lambda$, and where $\Gamma_{m\,n}^{\ r}$ and g^{mn} are independent variables of the gravitational field. Show that the action principle again gives the equations of general relativity theory, i.e., Equations (5.13), (5.14) as well as $\Gamma_{m\,n}^{\ r} = \frac{1}{2}g^{rs}(g_{sn,m} + g_{sm,n} - g_{mn,s})$.

Lie Differential and Derivative. A variational principle yields not only field equations, but also identities which follow from invariance properties of the action. The coordinate invariance of the action leads to the differential conservation laws of energy and momentum. In order to show this we shall now study the behavior of tensors under infinitesimal transformations.

Let ξ^i be a contravariant vector field, and consider the infinitesimal transformation

$$\bar{x}^i = x^i + \xi^i du. \tag{5.16}$$

This transformation can be viewed in two ways.

It can be viewed as a *passive* or coordinate *transformation,* where x^i and $\bar{x}^i$ label the same point P in different coordinate systems. A tensor field $T^{m\cdots}_{n\cdots}$ will then transform according to the equation

$$\bar{T}^{m\cdots}_{n\cdots}(\bar{x}^i) = \frac{\partial \bar{x}^m}{\partial x^r} \cdots \frac{\partial x^s}{\partial \bar{x}^n} \cdots T^{r\cdots}_{s\cdots}(x^i). \tag{5.17}$$

This is illustrated in Figure 19a.

It can also be viewed as an *active transformation,* where x^i and $\bar{x}^i$ label two different points P and $\bar{P}$ in the same coordinate system. This is illustrated in Figure 19b. The field $\bar{T}^{m\cdots}_{n\cdots}(\bar{x}^i)$, given by Equation (5.17), is then a new tensor field and is said to have

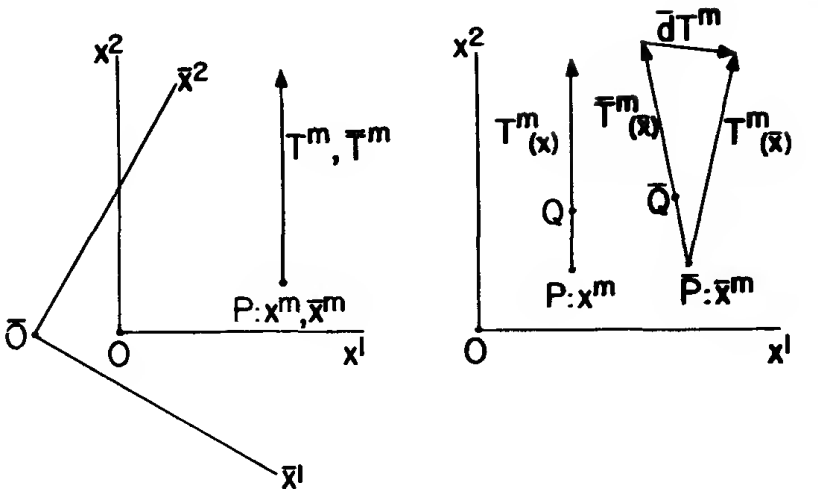


FIGURE 19

been obtained from $T^{m}_{n}{}^{\cdots}_{\cdots}(x^i)$ by letting the active infinitesimal transformation *drag* $T^{m}_{n}{}^{\cdots}_{\cdots}$ *with it.* If $T^m(x^i)$ is a multiple of an infinitesimal displacement PQ, then $\bar{T}^m(\bar{x}^i)$, is the same multiple of the infinitesimal displacement $\bar{P}\bar{Q}$. Similar constructions can be given for more complicated tensors or for other geometrical objects.

We retain the view of the active transformation, and write Equation (5.17) to first order in the infinitesimal du:

$$\begin{aligned}\bar{T}^{m}_{n}{}^{\cdots}_{\cdots}(\bar{x}^i) &= (\delta^m_r + \xi^m{}_{,r}du)\cdots(\delta^s_n - \xi^s{}_{,n}du)\cdots \\ &\quad \cdots (T^{r}_{s}{}^{\cdots}_{\cdots}(\bar{x}^i) - T^{r}_{s}{}^{\cdots}_{\cdots,i}\xi^i du) \\ &= T^{m}_{n}{}^{\cdots}_{\cdots}(\bar{x}^i) - T^{m}_{n}{}^{\cdots}_{\cdots,r}\xi^r du \\ &\quad + T^{r}_{n}{}^{\cdots}_{\cdots}\xi^m{}_{,r}du + \cdots - T^{m}_{r}{}^{\cdots}_{\cdots}\xi^r{}_{,n}du - \cdots.\end{aligned} \tag{5.18}$$

The *Lie differential* of $T^{m}_{n}{}^{\cdots}_{\cdots}$ is defined by

$$\begin{aligned}\bar{d}T^{m}_{n}{}^{\cdots}_{\cdots} &= T^{m}_{n}{}^{\cdots}_{\cdots}(\bar{x}^i) - \bar{T}^{m}_{n}{}^{\cdots}_{\cdots}(\bar{x}^i) \\ &= T^{m}_{n}{}^{\cdots}_{\cdots,r}\xi^r du - T^{r}_{n}{}^{\cdots}_{\cdots}\xi^m{}_{,r}du - \cdots \\ &\quad + T^{m}_{r}{}^{\cdots}_{\cdots}\xi^r{}_{,n}du + \cdots,\end{aligned} \tag{5.19}$$

and the *Lie derivative* by

$$\begin{aligned}\mathfrak{L}_\xi T^{m}_{n}{}^{\cdots}_{\cdots} &= \frac{\bar{d}}{du}T^{m}_{n}{}^{\cdots}_{\cdots} \\ &= T^{m}_{n}{}^{\cdots}_{\cdots,r}\xi^r - T^{r}_{n}{}^{\cdots}_{\cdots}\xi^m{}_{,r} - \cdots + T^{m}_{r}{}^{\cdots}_{\cdots}\xi^r{}_{,n} + \cdots.\end{aligned} \tag{5.20}$$

Note that the Lie derivative is nonmetrical.

Exercise. (1) Show that the Lie derivative of a tensor field is a tensor field of the same type.

(2) The Lie derivative obeys Leibnitz's rule.

(3) Lie differentiation commutes with contraction.

(4)

$$\pounds_\xi \delta^m{}_n = 0. \tag{5.21}$$

(5) In a Riemannian space,

$$\pounds_\xi T^{m\cdots}{}_{n\cdots} = T^{m\cdots}{}_{n\cdots;r}\xi^r - T^{r\cdots}{}_{n\cdots}\xi^m{}_{;r} - \cdots + T^{m\cdots}{}_{r\cdots}\xi^r{}_{;n} + \cdots. \tag{5.22}$$

(6) In a Riemannian space, Lie differentiation does not, in general, commute with raising and lowering of suffixes.

(7) Coordinates can be adapted to the vector field ξ^r, so that $\xi^1 = 1$, $\xi^2 = \xi^3 = \xi^4 = 0$. In these coordinates:

$$\pounds_\xi T^{m\cdots}{}_{n\cdots} = \partial T^{m\cdots}{}_{n\cdots}/\partial x^1. \tag{5.23}$$

Our results can be immediately applied to the study of symmetries of a Riemannian space. A Riemannian space is said to admit a *continuous one-parameter group of* symmetries or *motions* if, under a one-parameter group of active transformations, the distance between two neighboring points PQ is always the same as the distance between the image points $\overline{PQ}$. For the infinitesimal transformation (5.16) of the group, this means that $d\delta x^m = 0 \Rightarrow \bar{d}(g_{mn}\delta x^m \delta x^n) = 0$ or, equivalently, $\pounds_\xi \delta x^m = 0 \Rightarrow \pounds_\xi(g_{mn}\delta x^m \delta x^n) = 2g_{mn}\delta x^m \pounds_\xi(\delta x^n) + \delta x^m \delta x^n \pounds_\xi(g_{mn}) = 0$. Thus the condition for a continuous group of motions is the existence of a *Killing vector field* ξ^m which, by Equation (5.22), satisfies *Killing's equation*:

$$\pounds_\xi g_{mn} = 2\xi_{(m;n)} = 0. \tag{5.24}$$

The condition can also be stated in terms of the existence of adapted coordinates where $g_{mn,1} = 0$, so that the components of the metric tensor are independent of one of the coordinates.

Exercise. In Minkowski space-time, with rectangular Cartesians x, y, z, t, find the Killing vector fields for translations along the x- and t-axes, for rotations in the xy-plane and for Lorentz transformations in the xt-plane.

Conservation Laws. Consider the Lie differential

$$\bar{d}\int_{\tau(4)}(-g)^{1/2}Ld\tau_{(4)} \equiv \int_{\tau(4)}(-g)^{1/2}Ld\tau_{(4)} - \int_{\bar{\tau}(4)}(-\bar{g})^{1/2}\bar{L}d\bar{\tau}_{(4)}. \tag{5.25}$$

If L is an invariant field, then, adopting the passive view of the transformation (5.16), the two integrals on the right-hand side of Equation (5.25) are the same invariant integral evaluated in two different coordinate systems. Thus we have

$$\bar{d}\int_{\tau(4)}(-g)^{1/2}Ld\tau_{(4)} \equiv 0. \tag{5.26}$$

This result applies both to the gravitational action $I_{(G)}$ and to the field action $I_{(F)}$. Reverting to the active point of view, we may regard the Lie differential with respect to ξ^m as a special type of variation induced by the infinitesimal transformation (5.16). We now choose ξ^m so that ξ^m, $\xi^m{}_{,r}$ and $\xi^m{}_{,rs}$ vanish on the boundary $\tau_{(3)}$, although ξ^m remains arbitrary in the interior of $\tau_{(4)}$. Then $\bar{d}g^{mn}$, $\bar{d}g^{mn}{}_{,r}$ and $\bar{d}y_\mu$ vanish on $\tau_{(3)}$, and we can apply the results from the first section of this chapter.

We have

$$\begin{aligned} 0 \equiv \bar{d}I_{(G)} &\equiv \int_{\tau(4)}(-g)^{1/2}Q_{mn}\bar{d}g^{mn}d\tau_{(4)} \\ &\equiv du\int_{\tau(4)}(-g)^{1/2}Q_{mn}\,\mathfrak{L}_\xi(g^{mn})\,d\tau_{(4)} \end{aligned} \tag{5.27}$$

where

$$Q_{mn} \equiv G_{mn} + \Lambda g_{mn}. \tag{5.28}$$

By Equation (5.22),

$$\begin{aligned} (-g)^{1/2}Q_{mn}\,\mathfrak{L}_\xi g^{mn} &\equiv -2(-g)^{1/2}Q_{mn}\xi^{n;m} \\ &\equiv 2(-g)^{1/2}Q^m{}_{n;m}\xi^n - 2(-g)^{1/2}(Q^m{}_n\xi^n)_{;m} \\ &\equiv 2(-g)^{1/2}Q^m{}_{n;m}\xi^n - 2((-g)^{1/2}Q^m{}_n\xi^n)_{,m}. \end{aligned}$$

The integral of the last expression transforms into a boundary integral over $\tau_{(3)}$ and vanishes. Equation (5.27) therefore becomes

$$0 \equiv \bar{d} I_{(G)} \equiv 2du \int_{\tau(4)} (-g)^{1/2} Q^{m}{}_{n;m} \xi^{n} d\tau_{(4)}. \tag{5.29}$$

Since ξ^n is arbitrary at any point of $\tau_{(4)}$, it follows that

$$Q^{m}{}_{n;m} \equiv 0. \tag{5.30}$$

This is an identity, a consequence of the invariance of $I_{(G)}$, and completely independent of whether any field equations are satisfied or not. Identity signs have been used above to stress this.

In general relativity theory the identity (5.30) is equivalent to the fact that the divergence of the Einstein tensor vanishes, $G^{m}{}_{n;m} \equiv 0$. However, Equation (5.30) is more general. It remains valid for gravitational theories obtained from other Lagrangians, for example, for $L_{(G)} = R^{mnrs} R_{mnrs}$ when Q_{mn} and the gravitational field equations are of the fourth order.

The Lie derivative of the field action is, by Equation (5.9),

$$0 \equiv \bar{d} I_{(F)} \equiv du \int_{\tau(4)} (-g)^{1/2} [8\pi T_{mn} \pounds_{\xi} g^{mn} + L_{(F)}{}^{\mu} \pounds_{\xi} y_{\mu}] d\tau_{(4)}. \tag{5.31}$$

By Equation (5.20), $\pounds_{\xi} y_{\mu}$ is of the form

$$\pounds_{\xi} y_{\mu} \equiv y_{\mu,n} \xi^{n} - F_{\mu}{}^{\nu}{}_{n}{}^{m} y_{\nu} \xi^{n}{}_{,m}, \tag{5.32}$$

where $F_{\mu}{}^{\nu}{}_{n}{}^{m}$ are constants which are determined by the tensor character of y_{μ}. Proceeding as we did above for $I_{(G)}$, we obtain

$$\begin{aligned} 0 \equiv \bar{d} I_{(F)} \equiv du \int_{\tau(4)} [&16\pi (-g)^{1/2} T^{m}{}_{n;m} + L_{(F)}{}^{\mu} y_{\mu,n} \\ &+ ((-g)^{1/2} L_{(F)}{}^{\mu} F_{\mu}{}^{\nu}{}_{n}{}^{m} y_{\nu})_{,m}] \xi^{n} d\tau_{(4)}, \end{aligned} \tag{5.33}$$

and hence the identities

$$16\pi (-g)^{1/2} T^{m}{}_{n,m} + L_{(F)}{}^{\mu} y_{\mu,n} + ((-g)^{1/2} L_{(F)}{}^{\mu} F_{\mu}{}^{\nu}{}_{n}{}^{m} y_{\nu})_{,m} \equiv 0. \tag{5.34}$$

The conservation laws

$$T^{m}{}_{n,m} = 0 \tag{5.35}$$

now follow as a consequence of the field equations $L_{(F)}{}^{\mu} = 0$. They are, however, independent of the gravitational field equations. In the purely electromagnetic case, $E^{m}{}_{n;m} = 0$ is a consequence of Maxwell's equations.

This chapter has shown that a variational principle with an invariant action always gives field theories with some desirable

properties: (1) coordinate independent field equations, (2) a stress energy tensor, which (3) satisfies conservation laws.

6. **Equations of motion.** In theoretical physics, general relativity has the unique feature that the equations of motion of particles are not independent of the field equations. This is due, in part, to the nonlinearity of the gravitational field equations. In a linear field theory this property would be undesirable. We illustrate this, taking Maxwell's theory as an example. Let $\bar{F}_{mn}$ be a solution which has a single singularity along a world line $\bar{\Gamma}$ corresponding to a charge $\bar{e}$, and let $\bar{\bar{F}}_{mn}$ be a solution with a single singularity along $\bar{\bar{\Gamma}}$ corresponding to a charge $\bar{\bar{e}}$ (Figure 20). The superposition

$$F_{mn} = \bar{F}_{mn} + \bar{\bar{F}}_{mn} \tag{6.1}$$

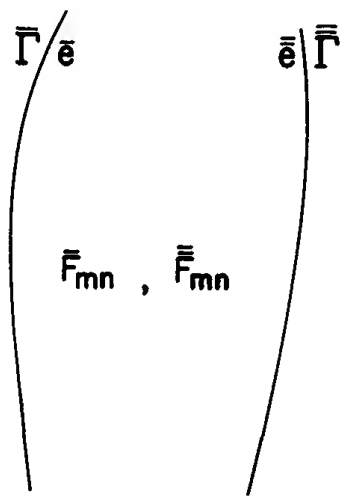


FIGURE 20

is then again a solution of Maxwell's equations, with charges $\bar{e}$ and $\bar{\bar{e}}$ moving along $\bar{\Gamma}$ and $\bar{\bar{\Gamma}}$, respectively. If the field equations determined the motion of the sources, then $\bar{e}$ would move exactly the same way whether $\bar{\bar{e}}$ were present or not; the charges could not interact through the field. This objection does not apply to a nonlinear field theory.

The quickest way to see a connection in general relativity between field equations and motion is to consider dust, whose stress energy tensor has the form

$$T^{mn} = \tilde{\rho} u^m u^n, \qquad u^m u_m = 1. \tag{6.2}$$

The gravitational field equations imply the differential conservation laws

$$0 = T^{mn}{}_{;n} = (\tilde{\rho}u^n)_{;n}u^m + \tilde{\rho}u^m{}_{;n}u^n. \tag{6.3}$$

Multiplying by u_m, we obtain

$$(\tilde{\rho}u^n)_{;n} = 0, \tag{6.4}$$

and hence

$$u^m{}_{;n}u^n = 0. \tag{6.5}$$

This last equation states that the curves with u^m as tangents are geodesics; the dust particles have geodesic world lines.

There are two main methods for dealing with the problem of motion.

The Einstein-Infeld-Hoffman method[14] applies to a finite number of slow particles interacting through a weak quasi-static gravitational field. In a scheme of successive approximations, the field equations are solved, and equations of motion are derived for the particles from the condition that the field equations of the next approximation be consistent.

The second method studies a single test particle moving in an arbitrarily strong gravitational field. This can be done by analysing either the vacuum gravitational field outside the particle,[15] or the internal structure of the test particle.[16] We shall discuss the last procedure, and give a simplified version of the work of Papapetrou and Taub. It may in fact be an oversimplified version, because it will be based on an important assumption which has never been proved.

Before returning to the problem of motion, we must learn some more tensor analysis.

Fermi Propagation. Let u^m be the unit tangent vector along a

---

[14] A. Einstein, L. Infeld and B. Hoffman, Ann. Math. 39 (1938), 65; 41 (1940), 455; Canad. J. Math. 1 (1949), 209; R. P. Kerr, Nuovo Cimento 13 (1959), 469, 492; 16 (1960), 26; L. Infeld and J. Plebanski, *Motion and relativity*, PWN, Warsaw, 1960; see also the review article by J. N. Goldberg in *Gravitation: an introduction to current research*, Wiley, New York, 1962.

[15] L. Infeld and A. Schild, Rev. Mod. Phys. 21 (1949), 408; J. R. Porter, Texas University Ph. D. Thesis, 1965.

[16] M. Mathisson, Acta Phys. Polon. 6 (1937), 163; A. Papapetrou, Proc. Roy. Soc. Ser A 209 (1951), 248; A. H. Taub, J. Math. Phys. 5 (1964), 112.

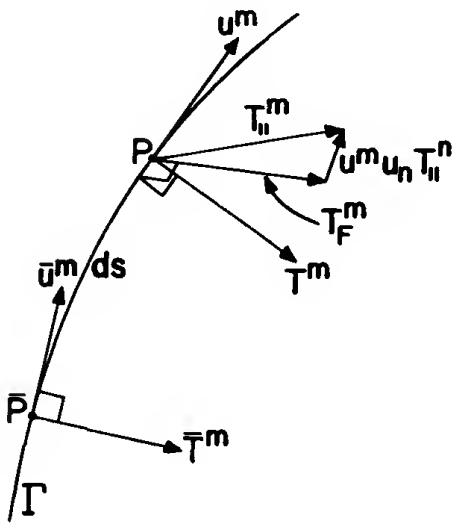


FIGURE 21

time-like curve Γ, and let T^m be a vector field along Γ and orthogonal to it:

$$u_m T^m = 0. \tag{6.6}$$

The vector $T_{\|}^m$ at P, obtained by parallel propagation of $\bar{T}^m$ at the neighboring point $\bar{P}$ of Γ (Figure 21), will in general not be orthogonal to Γ. The vector T_F^m is the orthogonal projection of $T_{\|}^m$:

$$T_F{}^m = P^m{}_n T_{\|}{}^n, \quad P^m{}_n = \delta^m{}_n - u^m u_n; \tag{6.7}$$

$T_F{}^m$ is said to be obtained by *Fermi propagation* of $\bar{T}^m$ from $\bar{P}$ to P. $DT^m = T^m - T_{\|}{}^m$ is called the *absolute differential* of T^m and $D_F T^m = T^m - T_F{}^m$ is the *Fermi differential*. By Equation (6.7), we have

$$\begin{aligned} D_F T^m &= DT^m + u^m u_n (T^n - DT^n) \\ &= DT^m - u^m u_n DT^n. \end{aligned} \tag{6.8}$$

The *Fermi derivative* $D_F T^m/ds = \overset{\circ}{T}{}^m$ is given by

$$\overset{\circ}{T}{}^m = \frac{DT^m}{ds} - u^m u_n \frac{DT^n}{ds}. \tag{6.9}$$

T^m is Fermi propagated along Γ if $\overset{\circ}{T}{}^m = 0$.

More generally, a tensor field $T_{m\cdots}^{n\cdots}$ is orthogonal to Γ if

$$T_{m\cdots}^{n\cdots} = P^r{}_m P^n{}_s \cdots T_{r\cdots}^{s\cdots}, \tag{6.10}$$

and its Fermi derivative is then defined by

$$\overset{\circ}{T}{}_{m\cdots}^{n\cdots} = P^{r}{}_{m} P^{n}{}_{s} \cdots \frac{DT_{r\cdots}^{s\cdots}}{ds}. \tag{6.11}$$

For a skew symmetric tensor $S^{mn} = -S^{nm}$ this is equivalent to

$$S^{mn} u_n = 0, \tag{6.12}$$

$$\overset{\circ}{S}{}^{mn} = \frac{DS^{mn}}{ds} - u^{m} u_{r} \frac{DS^{rn}}{ds} - u^{n} u_{r} \frac{DS^{mr}}{ds}. \tag{6.13}$$

Exercise. 1. Show that the Fermi derivative obeys Leibnitz's rule.

2. Show that the magnitude of a vector and the angle between two vectors remain constant under Fermi propagation.

Fermi Coordinates of the Second Kind. At one point 0 of the time-like curve Γ (Figure 22) choose a triad of orthonormal space-like vectors $e_{\alpha}{}^{m}$ orthogonal to Γ:

$$e_{\alpha}{}^{m} e_{\beta m} = -\delta_{\alpha\beta}, \qquad u_{m} e_{\alpha}{}^{m} = 0. \tag{6.14}$$

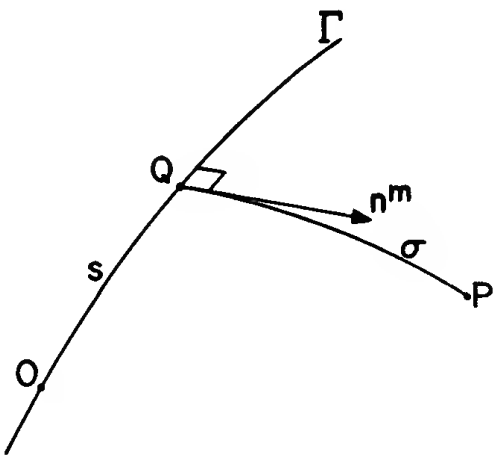


FIGURE 22

Fermi propagate the vectors $e_{\alpha}{}^{m}$ along Γ, so that Equations (6.14) hold at all points of Γ. Then define the new coordinates x'^{m} just as was done in Chapter 3 for Fermi coordinates of the first kind, except that the tetrad $e_{\alpha}{}^{m}$, $e_{4}{}^{m} = u^{m}$ now replaces the parallelly propagated tetrad used there. This is equivalent to defining:

$$(x'^{\mu})_{P} = n^{\mu}\sigma, \qquad (x'^{4})_{P} = s, \tag{6.15}$$

where s is the arc length OQ, σ the arc length of the geodesic QP, and n^m the unit tangent

$$n^m = n^\alpha e_\alpha{}^m, \qquad n^\alpha n^\alpha = 1 \tag{6.16}$$

to QP at Q.

Dropping primes, the following relations hold along Γ in Fermi coordinates of the second kind (exercise):

$$g_{mn} = \eta_{mn}, \tag{6.17}$$

$$\left\{ \begin{matrix} m \\ \alpha \ \beta \end{matrix} \right\} = 0, \left\{ \begin{matrix} \beta \\ \alpha \ 4 \end{matrix} \right\} = 0, \left\{ \begin{matrix} 4 \\ 4 \ 4 \end{matrix} \right\} = 0,$$

$$\left\{ \begin{matrix} \alpha \\ 4 \ 4 \end{matrix} \right\} = \left\{ \begin{matrix} 4 \\ 4 \ \alpha \end{matrix} \right\} = a^\alpha, \tag{6.18}$$

where the world acceleration of Γ is

$$\frac{Du^m}{ds} = (a^\mu, 0). \tag{6.19}$$

The curve Γ has the equation $x^\mu = 0$, so that $u^m = (0,0,0,1)$. The Fermi derivative along Γ of a tensor field $T_{\cdot\cdot}$ orthogonal to Γ is given by

$$\mathring{T}_{\cdot\cdot} = \dot{T}_{\cdot\cdot}, \tag{6.20}$$

where $\dot{} \equiv (d/dx^4)$. Fermi propagation of $T_{\cdot\cdot}$ reduces to the condition $T_{\cdot\cdot}$ = constant in Fermi coordinates of the second kind.

When $a^\mu = 0$, Γ is a geodesic and the coordinates become *geodesic Fermi coordinates*. In this case the distinction between Fermi coordinates of the first and second kind disappears.

Motion of a Test Particle. Consider an isolated system, i.e., a time-like world tube outside of which $T^{mn} = 0$ (Figure 23). We shall call this a test particle if the tube is thin and if the components of T^{mn} inside the tube are small.

Let Γ be a time-like world line and let x^m be Fermi coordinates of the second kind for Γ. We now make the *assumption* that Γ can be chosen, so that all along Γ

$$\int x^\alpha \zeta^{44} dv = 0. \tag{6.21}$$

Here, and throughout this chapter

$$\zeta^{mn} = (-g)^{1/2} T^{mn}, \tag{6.22}$$

the integrals are taken over sections $x^4 = \text{constant}$ of the tube, and $dv = dx^1dx^2dx^3$ is the 3-extension of a 3-cell with edges in the directions of the x^α-coordinate lines. This assumption has never been proved, but we shall accept it as valid, and from here on Γ is such that Equation (6.21) holds. Since ζ^{44} is nonnegative, Γ lies inside the convex hull of the world tube.

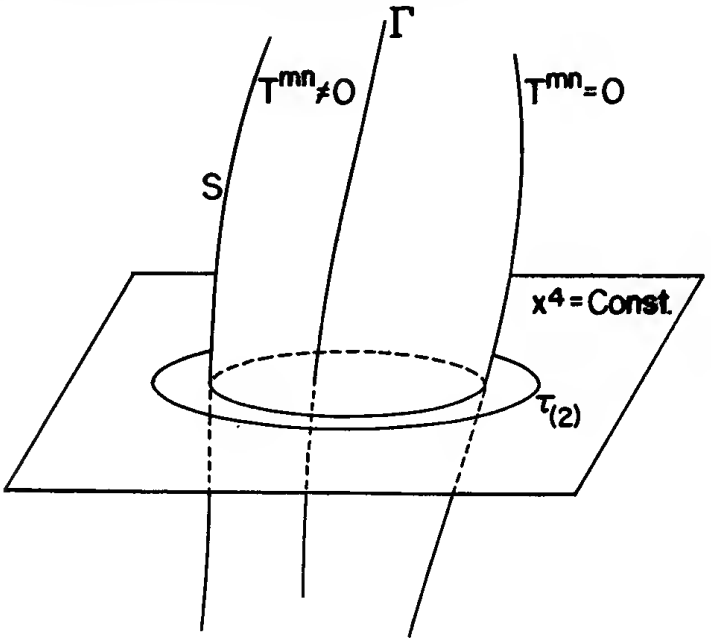


FIGURE 23

Consider the moments

$$M^{mn} = \int \zeta^{mn} dv, \tag{6.23}$$

$$M^{\alpha mn} = \int x^\alpha \zeta^{mn} dv, \tag{6.24}$$

$$M^{\alpha\beta mn} = \int x^\alpha x^\beta \zeta^{mn} dv, \cdots. \tag{6.25}$$

We call the test particle a *simple pole particle*, or a particle without spin if the first and higher moments are negligibly small:

$$M^{\alpha mn} \approx 0, \qquad M^{\alpha\beta mn} \approx 0, \cdots. \tag{6.26}$$

We call it a *pole-dipole particle*, or a particle with spin but no higher moments, if

$$M^{\alpha\beta mn} \approx 0, \cdots. \tag{6.27}$$

We shall limit ourselves here to the motion of a pole-dipole particle, although the method given can easily be extended to particles with higher multipole moments.

The conservation equations $T^{mn}{}_{;n} = 0$ are equivalent to (exercise):

$$\zeta^{mn}{}_{,n} = -\left\{ \begin{matrix} m \\ r \quad n \end{matrix} \right\} \zeta^{rn}. \tag{6.28}$$

From this follows

$$(x^\alpha \zeta^{mn})_{,n} = \zeta^{m\alpha} - \left\{ \begin{matrix} m \\ r \quad n \end{matrix} \right\} x^\alpha \zeta^{rn}, \tag{6.29}$$

$$(x^\alpha x^\beta \zeta^{mn})_{,n} = x^\alpha \zeta^{\beta m} + x^\beta \zeta^{\alpha m} - \left\{ \begin{matrix} m \\ r \quad n \end{matrix} \right\} x^\alpha x^\beta \zeta^{rn}. \tag{6.30}$$

We now integrate each of these equations over a section $x^4 =$ constant. The integrals of $\zeta^{m\nu}{}_{,\nu}$, $(x^\alpha \zeta^{m\nu})_{,\nu}$, $(x^\alpha x^\beta \zeta^{m\nu})_{,\nu}$ transform, by Green's theorem in three dimensions, into integrals over 2-surfaces $\tau_{(2)}$ (Figure 23) which surround the tube and therefore vanish, since $\zeta^{mn} = 0$ there. We also use the Taylor expansion

$$\left\{ \begin{matrix} m \\ r \quad n \end{matrix} \right\} = {}_0\left\{ \begin{matrix} m \\ r \quad n \end{matrix} \right\} + x^\alpha {}_0\left\{ \begin{matrix} m \\ r \quad n \end{matrix} \right\}_{,\alpha} + \cdots \tag{6.31}$$

in each section $x^4 =$ constant, where

$${}_0\left\{ \begin{matrix} m \\ r \quad n \end{matrix} \right\}, \quad {}_0\left\{ \begin{matrix} m \\ r \quad n \end{matrix} \right\}_{,\alpha}$$

indicates that the expressions are evaluated on Γ where $x^\mu = 0$. We then obtain, remembering Equation (6.27),

$$\dot{M}^{m4} = -{}_0\left\{ \begin{matrix} m \\ r \quad n \end{matrix} \right\} M^{rn} - {}_0\left\{ \begin{matrix} m \\ r \quad n \end{matrix} \right\}_{,\alpha} M^{\alpha rn}, \tag{6.32}$$

$$\dot{M}^{\alpha m4} = M^{\alpha m} - {}_0\left\{ \begin{matrix} m \\ r \quad n \end{matrix} \right\} M^{\alpha rn}, \tag{6.33}$$

$$M^{\alpha\beta m} + M^{\beta\alpha m} = 0, \tag{6.34}$$

where the dot denotes differentiation with respect to x^4.

The last equation tells us that $M^{\alpha\beta m}$ is skew symmetric in its first two suffixes. Also, by Equation (6.24), it is symmetric in the last two suffixes. Therefore

$$0 = (M^{\alpha\beta\gamma} + M^{\beta\alpha\gamma}) - (M^{\beta\gamma\alpha} + M^{\gamma\beta\alpha}) + (M^{\gamma\alpha\beta} + M^{\alpha\gamma\beta}) = 2M^{\alpha\beta\gamma}.$$

We thus have, by Equation (6.21),

$$M^{\alpha\beta\gamma} = 0, \qquad M^{\alpha 44} = 0. \tag{6.35}$$

The only nonvanishing component of $M^{\alpha mn}$ is

$$S^{\alpha\beta} = -S^{\beta\alpha} = 2M^{\alpha\beta 4} = 2M^{\alpha 4\beta}. \tag{6.36}$$

We call $S^{\alpha\beta}$ the *spin of the test particle.* We also call

$$m = M^{44} \tag{6.37}$$

the *mass of the test particle.*

Remembering that we are working in Fermi coordinates, Equations (6.32) and (6.33) become

$$\dot{M}^{\alpha 4} = -{}_{\circ}\left\{ {\alpha \atop 4\ 4} \right\} m - {}_{\circ}\left\{ {\alpha \atop \gamma\ 4} \right\}_{,\beta} S^{\beta\gamma}, \tag{6.38}$$

$$\dot{m} = -2\,{}_{\circ}\left\{ {4 \atop \alpha\ 4} \right\} M^{\alpha 4} - {}_{\circ}\left\{ {4 \atop \beta\ 4} \right\}_{,\alpha} S^{\alpha\beta}, \tag{6.39}$$

$$\dot{S}^{\alpha\beta} = 2M^{\alpha\beta}, \tag{6.40}$$

$$M^{\alpha 4} = {}_{\circ}\left\{ {4 \atop \beta\ 4} \right\} S^{\alpha\beta}. \tag{6.41}$$

The right-hand side of Equation (6.40) is symmetric in α and β, and the left-hand side is skew symmetric. Therefore $M^{\alpha\beta} = 0$ and

$$\dot{S}^{\alpha\beta} = 0. \tag{6.42}$$

It is easy to show that in Fermi coordinates

$${}_{\circ}\left\{ {4 \atop \beta\ 4} \right\}_{,\alpha} = \frac{1}{2}\left({}_{\circ}g_{44,\alpha\beta} - {}_{\circ}g_{44,\alpha}\,{}_{\circ}g_{44,\beta}\right)$$

is symmetric in α and β, so that

$${}_{\circ}\left\{ {4 \atop \beta\ 4} \right\}_{,\alpha} S^{\alpha\beta} = 0.$$

Also, by Equation (6.41),

$${}_{\circ}\left\{ {4 \atop \alpha\ 4} \right\} M^{\alpha 4} = {}_{\circ}\left\{ {4 \atop \alpha\ 4} \right\}\,{}_{\circ}\left\{ {4 \atop \beta\ 4} \right\} S^{\alpha\beta} = 0,$$

because $S^{\alpha\beta}$ is skew symmetric. Therefore Equation (6.39) reduces to

$$\dot{m} = 0. \tag{6.43}$$

The remaining equations, (6.38) and (6.41), give, upon elimination of $M^{\alpha 4}$:

$$(6.44)\quad m\,{}_{\circ}\left\{ {\alpha \atop 4\ 4} \right\} + \frac{d}{dx^4}\left({}_{\circ}\left\{ {4 \atop \beta\ 4} \right\} S^{\alpha\beta} \right) + {}_{\circ}\left\{ {\alpha \atop \gamma\ 4} \right\}_{,\beta} S^{\beta\gamma} = 0.$$

We introduce the *spin tensor* $S^{mn} = -S^{nm}$ of the particle as a skew symmetric tensor field defined along Γ and orthogonal to it,

$$(6.45)\qquad S^{mn}u_n = 0,$$

so that it reduces in Fermi coordinates to $S^{m4} = 0$ and $S^{\alpha\beta}$ given by Equation (6.36).

In Fermi coordinates we have

$$\frac{D}{ds}(mu^m) = \left(m\,{}_{\circ}\left\{ {\mu \atop 4\ 4} \right\}, 0 \right),$$

$$\frac{D}{ds}\left(-\frac{Du_n}{ds} S^{mn} \right) = \left(\frac{d}{dx^4}\left({}_{\circ}\left\{ {4 \atop \beta\ 4} \right\} S^{\mu\beta} \right), 0 \right),$$

$$\frac{1}{2} R^m{}_{nrs} u^n S^{rs} = \frac{1}{2} R^m{}_{4\rho\sigma} S^{\rho\sigma}$$

$$= {}_{\circ}\left\{ {m \atop 4\ \sigma} \right\}_{,\rho} S^{\rho\sigma} + {}_{\circ}\left\{ {4 \atop 4\ \sigma} \right\} {}_{\circ}\left\{ {m \atop 4\ \rho} \right\} S^{\rho\sigma}$$

$$= \left({}_{\circ}\left\{ {\mu \atop 4\ \sigma} \right\}_{,\rho} S^{\rho\sigma}, 0 \right).$$

Thus equations (6.44), (6.42), (6.45) can be written in tensor form

$$(6.46)\qquad \frac{D}{ds}\left(mu^m - \frac{Du_n}{ds} S^{mn} \right) + \frac{1}{2} R^m{}_{nrs} u^n S^{rs} = 0,$$

$$(6.47)\qquad \overset{\circ}{S}{}^{mn} = 0, \qquad S^{mn}u_n = 0.$$

These are the equations of motion of a spinning test particle.[17] The condition dm/ds is a consequence of these equations (exercise) and need not be assumed independently.

Exercise. Show that the *spin vector* S^m, defined by

$$(6.48)\qquad S^m = -\tfrac{1}{2}\eta^{mrst} S_{rs} u_t,$$

[17] F. A. E. Pirani, Acta Phys. Polon. **15** (1956), 389.

is orthogonal to and Fermi propagated along Γ; also that

$$S^{rs} = \eta^{rsmt} S_m u_t. \tag{6.49}$$

In Fermi coordinates of the second kind both equations reduce to

$$(S^1, S^2, S^3) = (S^{23}, S^{31}, S^{12}), \quad S^4 = 0, \quad S^{m4} = 0. \tag{6.50}$$

From the Fermi propagation of the spin, Equation (6.47), it follows that the magnitude of the spin

$$S = (-S^m S_m)^{1/2} = \left(\frac{1}{2} S^{mn} S_{mn}\right)^{1/2} \tag{6.51}$$

remains constant (exercise). However, the direction of the spin axis S^m moves. A proposed experiment to verify general relativity theory is based on this. It consists in measuring the precession of a gyroscope in an earth satellite.[18] (See Schiff's lectures in this volume.)

For a *simple pole particle* without spin, $S^{mn} = 0$ and the equation of motion (6.46) reduces to that of a geodesic,

$$\frac{Du^m}{ds} = 0, \tag{6.52}$$

in agreement with the principle of equivalence.

It is not surprising that a particle with spin deviates from geodesic motion, and that there is a driving term in Equation (6.46) which involves the curvature tensor. We saw in special relativity theory that an isolated system with spin has a minimum spatial extension given by Equation (2.50). We can expect the same thing to be true in general relativity theory: A test particle with spin has a finite extension (of minimum magnitude determined by the ratio S/m) and its motion is therefore affected by the gradient of the gravitational field strength.

Our derivation of the equations of motion of a test particle was based on the unproved assumption that a curve Γ exists in general relativity theory which has the properties of the local center of mass discussed in Chapter 2. In special relativity theory, the local center of mass is not unique, and there exists a particularly simple one, the relativistic center of mass. We may ask if something

[18] G. E. Pugh, Weapons System Evaluation Group Research Memorandum, No. 11, Washington, 1959; L. I. Schiff, Proc. Nat. Acad. Sci. U.S.A. 46 (1960), 871.

similar is true in general relativity theory. We shall conclude this chapter by stating some open questions. Any answers to these questions would, I believe, be an important contribution to general relativity theory.

Three Open Questions. 1. Given any isolated system, does there always exist a curve Γ so that Equation (6.21) holds in Fermi coordinates of the second kind?

2. Does such a curve exist in the approximation of a test particle?

3. If the answer to either of these questions is affirmative, is Γ unique? If Γ is not unique, can it be chosen so that the translational equations of motion (6.46) simplify to

$$\frac{Dmu^m}{ds} + \frac{1}{2} R^m{}_{nrs} u^n S^{rs} = 0? \tag{6.53}$$

7. **Jump conditions and Cauchy problem.** In this chapter we shall study the structure of the gravitational field equations. For this purpose it is necessary first to discuss the smoothness of space-time.

In general, a space cannot be covered by a single well-behaved coordinate system. An example of this is provided by the 2-dimensional surface of a sphere in Euclidean 3-space. If a single well-behaved system of coordinates x^1, x^2 covered the 2-sphere, then $(T^1, T^2) = (1, 0)$ would be a continuous vector field which vanishes nowhere, and this contradicts a well-known fixed point theorem.

We follow Lichnerowicz,[19] and assume that space-time can be broken up into overlapping domains, with a system of coordinates x^m in each domain. Where two such domains or *coordinate patches* overlap, we assume that the coordinate transformation $x'^m(x^1, x^2, x^3, x^4)$ satisfies these criteria: (1) The transformation functions have continuous fourth derivatives except on some hypersurfaces where they have continuous second derivatives; the third and fourth derivatives have definite limits as such a hypersurface is approached from either side but may be discontinuous across it; (2) starting with one of the two coordinate systems, a transformation with continuous second derivatives exists which reduces the equation of a hypersurface of discontinuity to the form $x^4 = 0$; (3) the Jacobian determinant $|\partial x'/\partial x|$ is nowhere zero. We say that the coordinate

[19] A. Lichnerowicz, *Théories relativistes de la gravitation et de l'Électromagnétisme*, Masson et Cie., Paris, 1955.

transformations are $[C^2; C^4$ by pieces$]$. The properties just outlined make space-time a *differentiable manifold* of class $[C^2; C^4$ by pieces$]$. A set of coordinate patches which covers a differentiable manifold is called an *atlas*. Two atlases are said to be *equivalent* if the union of the two sets of coordinate patches is again an atlas. Equivalent atlases provide different coordinatizations of the same differentiable manifold. In fact, a differentiable manifold can be defined abstractly as the collection of all equivalent atlases.

We now introduce the metric g_{mn} of space-time, with signature $(- - - +)$ and with the following conditions of smoothness: In each coordinate patch, the functions $g_{mn}(x^1, x^2, x^3, x^4)$ are $[C^1; C^3$ by pieces$]$, i.e., they have continuous third derivatives except on some hypersurfaces, where they have continuous first derivatives, and where the second and third derivatives have definite limits as the hypersurface is approached from either side but may be discontinuous across it. Where two coordinate patches overlap, the components of the metric tensor are related by the tensor transformation law, which involves first derivatives of the coordinate transformation. It is because we wish the metric tensor to have smoothness $[C^1; C^3$ by pieces$]$, that we must insist on coordinate transformations with smoothness $[C^2; C^4$ by pieces$]$, so that their first derivatives are $[C^1; C^3$ by pieces$]$.

The assumptions we have made about the smoothness of g_{mn} are similar to those which apply to the Newtonian gravitational potential V when the volume density of matter is assumed not to be infinite.

Discontinuities Across a Hypersurface. Let the hypersurface S have the equation $x^4 = 0$. We denote by $[C]$ quantities which are continuous across S. Then

$$g_{mn} = [C], \quad g_{mn,r} = [C], \quad g_{mn,r\alpha} = [C], \tag{7.1}$$

but $g_{mn,44}$ may be discontinuous.

From Table I, we have for the Riemann tensor

$$R_{rsmn} = \tfrac{1}{2}\,(g_{rn,sm} + g_{sm,rn} - g_{rm,sn} - g_{sn,rm}) + [C]. \tag{7.2}$$

Therefore

$$R_{\rho\sigma\mu\nu} = [C], \quad R_{\rho\sigma\mu 4} = [C], \quad R_{\rho 4\mu 4} = -\tfrac{1}{2}\, g_{\rho\mu,44} + [C], \tag{7.3}$$

and this determines all independent components of the curvature tensor.

It is interesting that the discontinuity of the curvature tensor involves $g_{\rho\mu,44}$ but does not involve $g_{m4,44}$. In fact, a discontinuity

$$\Delta g_{m4,44}(x^1, x^2, x^3) = (g_{m4,44})_{x4\to 0+} - (g_{m4,44})_{x4\to 0-} \tag{7.4}$$

is unimportant and can always be wiped out by the coordinate transformation (exercise):

$$\begin{aligned} x'^m &= x^m + \frac{1}{6}(x^4)^3 A^m, \qquad x^4 > 0, \\ x'^m &= x^m, \qquad\qquad\qquad x^4 < 0, \end{aligned} \tag{7.5}$$

where A^m is determined by

$$A^m = g^{mn}A_n, \qquad A_\mu = \Delta g_{\mu 4,44}, \qquad A_4 = \tfrac{1}{2}\Delta g_{44,44}. \tag{7.6}$$

This transformation preserves the continuity relations (7.1).

From Equations (7.3), a simple calculation gives

$$G^4{}_m = [C], \tag{7.7}$$

$$R_{\rho\sigma} = \frac{1}{2} g^{44} g_{\rho\sigma,44} + [C]. \tag{7.8}$$

For a general hypersurface S with equation

$$\phi(x^1, x^2, x^3, x^4) = 0, \tag{7.9}$$

$d\phi = \phi_{,n}dx^n = 0$ for any displacement dx^n within the hypersurface. Thus

$$n_n = \phi_{,n} \tag{7.10}$$

is a normal to S. When the equation of S is $x^4 = 0$, we have

$$n_n = (0, 0, 0, 1). \tag{7.11}$$

We can therefore write Equation (7.7) in the form

$$n_n G^n{}_m = [C]. \tag{7.12}$$

This is a tensor relation and is therefore valid in general coordinates where the hypersurface of discontinuity has the equation (7.9).

Junction Conditions at Matter-Vacuum Interface. Consider a hypersurface S, with Equation (7.9), which separates matter ($T^{mn} \neq 0$) from vacuum ($T^{mn} = 0$). The boundary of the world

tube of an isolated system (Figure 23) is an example of such a hypersurface. The field equations $G^{mn} = -8\pi T^{mn}$ and the continuity condition (7.12) give

$$T^{mn}n_n = 0 \tag{7.13}$$

on S. Also, by hypothesis

$$g_{mn} = [C], \quad g_{mn,r} = [C] \tag{7.14}$$

across S. These are the junction conditions at an interface between matter and vacuum.

For an interface across which T^{mn} has a discontinuity ΔT^{mn}, without T^{mn} being zero on one side, Equation (7.14) holds and Equation (7.13) is replaced by

$$\Delta T^{mn}n_n = 0. \tag{7.15}$$

Let us examine the junction condition (7.13) more closely at one point P of the interface S. At one point P we can write T^{mn} as we did in special relativity theory (discussion following Equations (2.25) and (2.26)), in terms of the proper world velocity u^m, the proper density $\tilde{\rho}$ and the proper stress $\tilde{p}^{mn}$:

$$T^{mn} = \tilde{\rho}u^m u^n + \tilde{p}^{mn}, \quad u^m u_m = 1, \quad \tilde{p}^{mn}u_n = 0. \tag{7.16}$$

Equation (7.13) becomes

$$\tilde{\rho}u^m u^n n_n + \tilde{p}^{mn}n_n = 0.$$

Multiplication by u_m reduces this to

$$u^n n_n = 0, \tag{7.17}$$

$$\tilde{p}^{mn}n_n = 0. \tag{7.18}$$

The first of these equations states that the streamlines of the fluid lie in the hypersurface S; this agrees with a well-known boundary condition in classical continuum mechanics. The second Equation (7.18) is best examined in the proper frame, a coordinate system such that at P

$$g_{mn} = \eta_{mn}, \qquad u^m = (0, 0, 0, 1). \tag{7.19}$$

Then

$$n^m = (n^\mu, 0), \tag{7.20}$$

$$\tilde{p}^{mn} = \begin{pmatrix} \tilde{p}^{\mu\nu}, 0 \\ 0 \quad , 0 \end{pmatrix} . \tag{7.21}$$

Equation (7.18) now reduces to the other boundary condition of classical continuum mechanics, that the surface traction be zero:

$$\tilde{p}^{\mu\nu} n^{\nu} = 0. \tag{7.22}$$

Discontinuity of the Vacuum Riemann Tensor. We have seen that the only nontrivial discontinuity of $g_{mn,rs}$ across a hypersurface S occurs when the Riemann tensor is discontinuous. In special coordinates where S has the equation $x^4 = 0$ we have, by Equation (7.3),

$$\Delta R_{\rho 4 \mu 4} = -\frac{1}{2} \Delta g_{\rho\mu,44}, \tag{7.23}$$

and

$$\Delta R_{\rho\sigma\mu\nu} = 0, \qquad \Delta R_{\rho\sigma\mu 4} = 0, \tag{7.24}$$

where for any function $f(x^1, x^2, x^3, x^4)$, $\Delta f = (f)_{x4 \to 0+} - (f)_{x4 \to 0-}$.

We now assume that the vacuum gravitational field equations $R_{mn} = 0$ are satisfied. They hold on both sides of S, so that $\Delta R_{mn} = 0$, or

$$\Delta R_{\rho\mu} = -g^{44} \Delta R_{\rho 4 \mu 4} = 0, \tag{7.25}$$

$$\Delta R_{\rho 4} = g^{4\mu} \Delta R_{\rho 4 \mu 4} = 0, \tag{7.26}$$

$$\Delta R_{44} = -g^{\rho\mu} \Delta R_{\rho 4 \mu 4} = 0. \tag{7.27}$$

If the Riemann tensor is discontinuous across S, then $\Delta R_{\rho 4 \mu 4}$ are not all zero, and Equation (7.25) implies

$$g^{44} = 0. \tag{7.28}$$

In general coordinates, where S has the equation $\phi = 0$, this becomes

$$g^{mn} n_m n_n = 0, \qquad n_m = \phi_{,m}. \tag{7.29}$$

Such a hypersurface, whose normal is null and lies in the hypersurface, is called a *null hypersurface*. At each of its points it is tangent to the infinitesimal null cone along a null direction n^m, and it has no other tangential null directions (Figure 24). We say that the vacuum field equations $R_{mn} = 0$ are *hyperbolic*, and that the null hypersurfaces, across which nontrivial discontinuities of $g_{mn,rs}$

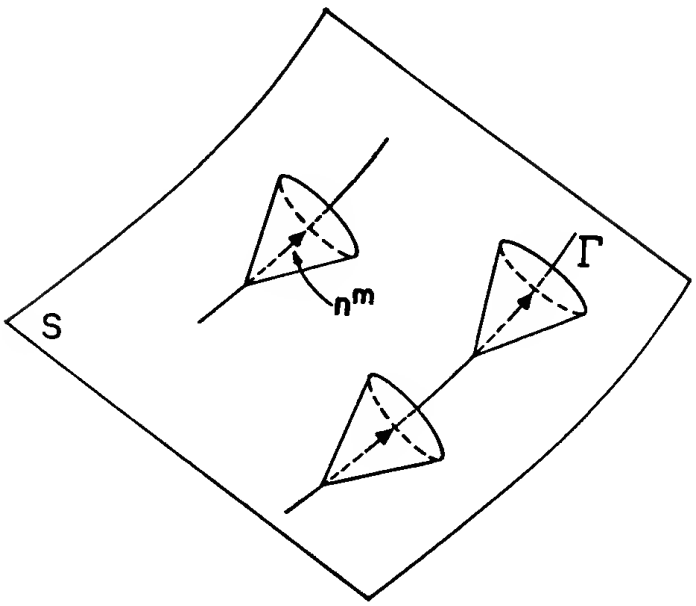


FIGURE 24

are possible, are the *characteristic hypersurfaces* of the field equations. We shall see shortly that an alternate definition of characteristics is possible, as hypersurfaces on which the Cauchy initial value problem is not well posed.

The discontinuity of the vacuum Riemann tensor is subject to restrictions which are given by Equations (7.24), (7.26), (7.27) in special coordinates. In general coordinates, the first two of these are easily seen to become (exercise):

$$\Delta R_{rs[mn}n_{t]} = 0, \quad \Delta R_{rsmn}n^n = 0. \tag{7.30}$$

In Chapter 10 we shall discuss the algebraic classification of vacuum Riemann tensors. Tensors ΔR_{rsmn} which satisfy Equations (7.30) will be called *null.*

Exercise. Integrate

$$((-g)^{1/2}F^{mn})_{,n} = 4\pi(-g)^{1/2}J^m \quad \text{and} \quad ((-g)^{1/2}F^{*mn})_{,n} = 0$$

over the small thin cylinder shown in Figure 25 which tightly envelops an element of a hypersurface S. Apply Green's theorem and take the limit as the height ϵ of the cylinder tends to zero. Deduce that a discontinuity ΔF^{mn} of the electromagnetic field across a hypersurface $\phi = 0$ must satisfy the conditions

(7.31) $\Delta F^{mn} n_n = 0, \quad \Delta F^{*mn} n_n = 0 \Leftrightarrow \Delta F_{[mn} n_{r]} = 0, \quad n_r = \phi_{,r}.$

Show that these equations imply that the characteristics of Maxwell's equations are the null hypersurfaces

(7.32) $$n^r n_r = 0.$$

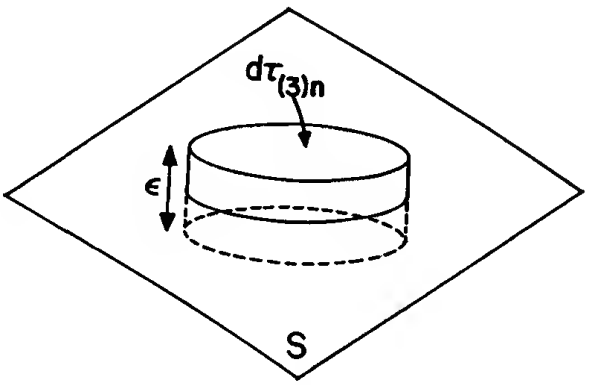


FIGURE 25

Characteristics and Bicharacteristics. The characteristics of the gravitational vacuum equations (and of Maxwell's equations) are the null hypersurfaces. The null geodesics are called *bicharacteristics* because null hypersurfaces contain these curves and can be constructed from them. This will now be shown.

In a null hypersurface S with equation $\phi = 0$, consider the null curves which are tangent to the normals $n_r = \phi_{,r}$:

(7.33) $$\frac{dx^r}{du} = n^r.$$

There is one such null curve through each point of S. We first imbed S in a family of null hypersurfaces with equations $\phi =$ constant. Then along the curve

(7.34) $$\frac{Dn_r}{du} = n^s n_{r;s} = n^s \phi_{;rs} = n^s \phi_{;sr}$$
$$= \frac{1}{2} (n^s n_s)_{,r} = 0,$$

since $n^s n_s = 0$. Thus the curves (7.33) are null geodesics, with u as affine parameter.

The most general null hypersurface can be constructed as follows.

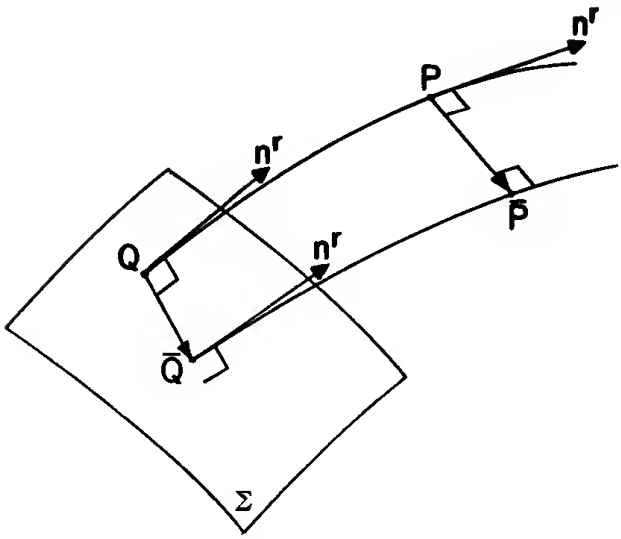


FIGURE 26

Start with a 2-surface Σ which is space-like, i.e., which at each point has exactly two null directions orthogonal to Σ. Choose one of these directions at one point of Σ and then, by continuity, the corresponding null direction n_r at all other points of Σ (Figure 26). Through each point of Σ draw the null geodesic with n_r as tangent. These null geodesics form a three-dimensional hypersurface S.

Neighboring null geodesics used in this construction have orthogonal connecting vectors, since on the initial surface Σ the connecting vectors are orthogonal (cf., exercise following Equation (4.21)). At a general point P of S, the tangent vector n^r to the null geodesic through P is orthogonal to itself and also to two independent connecting vectors from P to neighboring null geodesics. Therefore n^r is orthogonal to S and S is null.

Conversely, if S is a null hypersurface, choose a space-like 2-surface Σ in S. Our construction will now reproduce S. Therefore the most general null hypersurface can be obtained by the construction.

Cauchy Problem for Vacuum.[20] The Cauchy problem for second order partial differential equations consists in assigning on a hyper-

[20] A more detailed treatment of the Cauchy problem is found in A. Lichnerowicz, op. cit. The most rigorous discussion has been given by Y. Bruhat, J. Rat. Mech. Anal. 5, 951; *Gravitation, an introduction to current research,* J. Wiley and Sons, New York, 1962.

surface S the unknown functions and their first derivatives, and to determine, as far as possible, the solution of the differential equations in the neighborhood of S from the initial data.

We shall study this problem for the gravitational equations in vacuum:

$$R_{mn} = 0. \tag{7.35}$$

Let the hypersurface S have the equation $x^4 = \text{constant}$. The initial data on S consist of the functions

$$g_{mn}, \qquad g_{mn,4} \tag{7.36}$$

of the three variables x^1, x^2, x^3. By differentiation within S, the initial data determine also

$$g_{mn,\alpha}, \qquad g_{mn,\alpha\beta}, \qquad g_{mn,4\alpha} \tag{7.37}$$

on S. The quantities (7.36) and (7.37) are exactly those denoted by $[C]$ in Equation (7.1). We can therefore take over the results of Equation (7.3) to (7.8), reinterpreting $[C]$ to mean "expressions determined by the Cauchy initial data."

We see that the vacuum equations (7.35) cannot determine the four second derivatives $g_{m4,44}$ on S. They do determine the six second derivatives $g_{\rho\sigma,44}$, unless $g^{44} = 0$. If $g^{44} = 0$, only four combinations of $g_{\rho\sigma,44}$ are determined on S, which is then a null or characteristic hypersurface. We say that the Cauchy problem is not well posed on such a hypersurface.

From here on, we assume that S is space-like, so that $g^{44} > 0$. It is then easy to see (exercise) that

$$0 = G^4{}_m = [C], \tag{7.38}$$

$$0 = R_{\rho\sigma} = \frac{1}{2} g^{44} g_{\rho\sigma,44} + [C] \tag{7.39}$$

are independent combinations of the field equations (7.35) and therefore equivalent to them. The Cauchy problem is both underdetermined and overdetermined.

It is underdetermined because the gravitational field equations do not give $g_{m4,44}$ on S in terms of the initial data. This behavior is to be expected. The coordinates off the surface S can be changed freely, and we therefore cannot obtain unique functions $g_{mn}(x^1, x^2, x^3, x^4)$ in terms of initial data. The quantities $g_{m4,44}$ can be chosen arbitrarily,

different choices corresponding to different coordinate systems. The different solutions will, however, describe the same space-time geometry. In the next section we shall work with Gaussian normal coordinates, which are sufficiently rigid to eliminate the coordinate ambiguities. In these coordinates $g_{\mu 4} = 0$ and $g_{44} = 1$ throughout space-time, so that $g_{m4,44} = 0$.

The Cauchy problem is overdetermined, because the initial data cannot be given arbitrarily on S but must satisfy the *constraint equations* (7.38), $G^4{}_m = 0$. The identity $G^n{}_{m;n} = 0$ may be written

$$G^4{}_{m,4} = -G^\nu{}_{m,\nu} - \left\{ \begin{matrix} n \\ r \quad n \end{matrix} \right\} G^r{}_m + \left\{ \begin{matrix} r \\ m \quad n \end{matrix} \right\} G^n{}_r. \tag{7.40}$$

If the field equations are satisfied on S, then $G^n{}_m = 0$ and

$$G^4{}_{m,4} = 0 \tag{7.41}$$

on S.

We can now visualize the solution of the Cauchy problem as follows:

1. On the initial space-like surface S $(x^4 = \bar{x}^4 = \text{constant})$, $g_{mn}(x^1, x^2, x^3)$ and $g_{mn,4}(x^1, x^2, x^3)$ are given satisfying the four constraints $G^4{}_m = 0$. The four functions $g_{m4,44}(x^1, x^2, x^3)$ are chosen arbitrarily. The six functions $g_{\rho\sigma,44}(x^1, x^2, x^3)$ are determined by the field equations $R_{\rho\sigma} = 0$.

2. The quantities g_{mn}, $g_{mn,4}$ and $g_{mn,44}$ now determine $\bar{g}_{mn}(x^1, x^2, x^3)$, $\bar{g}_{mn,4}(x^1, x^2, x^3)$ on the neighboring hypersurface $\bar{S}(x^4 = \bar{x}^4 + d\bar{x}^4)$:

$$\bar{g}_{mn} = g_{mn} + g_{mn,4} d\bar{x}^4, \tag{7.42}$$

$$\bar{g}_{mn,4} = g_{mn,4} + g_{mn,44} d\bar{x}^4. \tag{7.43}$$

This constitutes initial data on the new hypersurface. The constraints are automatically satisfied on $\bar{S}$, since

$$\bar{G}^4{}_m = G^4{}_m + G^4{}_{m,4} d\bar{x}^4 = 0, \tag{7.44}$$

by Equation (7.41). We again choose $\bar{g}_{m4,44}$ arbitrarily on S, and let the field equations $R_{\rho\sigma} = 0$ determine $\bar{g}_{\rho\sigma,44}$.

3. We proceed in this manner from one neighboring hypersurface to the next.

This plausibility argument shows that the Cauchy problem gives a unique solution of the field equations, if the constraints

$G^4{}_m = 0$ are satisfied just once on the initial surface, and if $g_{m4,44}$ have been chosen throughout space-time.

Gaussian Normal Coordinates. Starting with a space-like hypersurface S, we draw through each of its points Q the geodesic QP orthogonal to S, and we denote the arc length QP by s (Figure 27). The points P for which s has the same value form a hypersurface which is again orthogonal to the geodesics QP. This follows from the theorem proved in Chapter 4, that if two neighboring time-like geodesics QP, $\bar{Q}\bar{P}$ have an orthogonal connecting vector $Q\bar{Q}$, and if the arc lengths QP and $\bar{Q}\bar{P}$ are equal, then the connecting vector $P\bar{P}$ is also orthogonal. The construction has provided us with a family of hypersurfaces and a congruence of orthogonal geodesics.

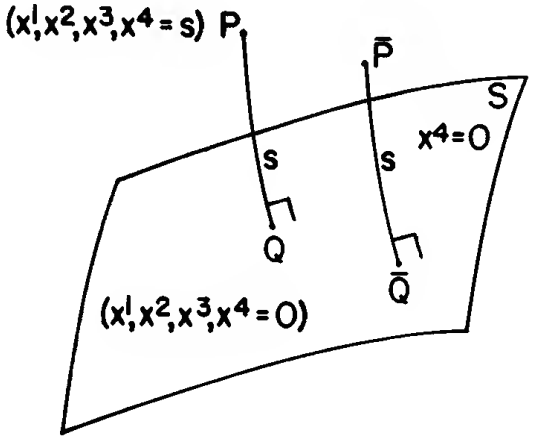


FIGURE 27

We now introduce Gaussian normal coordinates as follows. The points Q of the initial surface S are labelled by three parameters x^1, x^2, x^3, and we put

$$(x^4)_Q = 0. \tag{7.45}$$

For a general point P on the geodesic QP, we define

$$(x^\mu)_P = (x^\mu)_Q, \qquad (x^4)_P = s. \tag{7.46}$$

In these coordinates the family of hypersurfaces has the equations $x^4 =$ constant, the congruence of normal geodesics the equations $x^\mu =$ constant.

Gaussian normal coordinates are characterized by the property that

$$g_{\mu 4} = 0, \qquad g_{44} = 1 \tag{7.47}$$

throughout the domain of space-time where the coordinates are well behaved. From this we immediately deduce

$$g^{\mu 4} = 0, \qquad g^{44} = 1, \qquad g^{\mu\sigma} g_{\nu\sigma} = \delta^{\mu}{}_{\nu}, \tag{7.48}$$

and

$$\begin{Bmatrix} \rho \\ 4 \ \sigma \end{Bmatrix} = \frac{1}{2} g^{\rho\mu} g_{\mu\sigma,4}, \quad \begin{Bmatrix} 4 \\ \rho \ \sigma \end{Bmatrix} = -\frac{1}{2} g_{\rho\sigma,4},$$
$$\begin{Bmatrix} 4 \\ 4 \ \rho \end{Bmatrix} = 0, \quad \begin{Bmatrix} \rho \\ 4 \ 4 \end{Bmatrix} = 0, \quad \begin{Bmatrix} 4 \\ 4 \ 4 \end{Bmatrix} = 0. \tag{7.49}$$

The components $g_{\mu\nu}$ of the metric tensor g_{mn} determine the distances between neighboring points on the same hypersurface S with equation $x^4 =$ constant. We call $g_{\mu\nu}$ the *first fundamental form* of S. It describes the intrinsic or inner geometry of S, $g_{\mu\nu}$ and $g^{\mu\nu}$ being the submetric of the 3-dimensional subspace S. The submetric transforms like a tensor if x^4 is left unchanged and the coordinates x^1, x^2, x^3 are transformed among themselves. The Christoffel symbols

$$\begin{Bmatrix} \mu \\ \rho \ \sigma \end{Bmatrix}$$

formed from the submetric are identical with the corresponding components of

$$\begin{Bmatrix} m \\ r \ s \end{Bmatrix}$$

in the 4-dimensional parent space. We shall denote the Riemann and Ricci tensors and the curvature invariant of the subspace S by $\overline{R}_{\rho\sigma\mu\nu}$, $\overline{R}_{\rho\sigma}$, $\overline{R}$, respectively. They are defined in the usual way by the submetric $g_{\mu\nu}$.

Let n_m be the unit normal of a hypersurface S, $x^4 =$ constant, so that $n_m n^m = 1$. The covariant derivatives

$$K_{m\nu} = n_{m;\nu} \tag{7.50}$$

in directions tangential to S determine the rate at which the normal turns. They describe an extrinsic or outer curvature of S which depends on the particular way in which the subspace S is imbedded in the parent space. We have

$$n^m K_{m\nu} = \frac{1}{2}(n^m n_m)_{,\nu} = 0. \tag{7.51}$$

In Gaussian normal coordinates, $K_{4\nu} = 0$ and

$$K_{\mu\nu} = \frac{1}{2} g_{\mu\nu,4}. \tag{7.52}$$

We call $K_{\mu\nu}$ the *second fundamental form* of S.

Using Gaussian normal coordinates, we shall now state the Cauchy problem of the vacuum gravitational field equations in terms of the geometrically significant quantities $g_{\mu\nu}$ and $K_{\mu\nu}$. The quantities constitute the Cauchy initial data. The four constraint equations are

$$G^4_{\ \mu} = K^{\rho}_{\ \rho,\mu} - K^{\rho}_{\ \mu;\rho} = 0, \tag{7.53}$$

$$G^4_{\ 4} = -\frac{1}{2}\bar{R} + \frac{1}{2} K_{\rho\sigma}K^{\rho\sigma} - \frac{1}{2} K_{\rho}^{\ \rho} K_{\sigma}^{\ \sigma} = 0. \tag{7.54}$$

The other six field equations are

$$R_{\mu\nu} = K_{\mu\nu,4} + \bar{R}_{\mu\nu} - 2K_{\mu}^{\ \sigma}K_{\nu\sigma} + K_{\mu\nu}K_{\sigma}^{\ \sigma} = 0. \tag{7.55}$$

They contain the time derivatives $K_{\mu\nu,4}$ and determine the evolution of the field into the future. In Equation (7.53), $K^{\rho}_{\ \mu;\rho}$ is a covariant derivative with respect to the submetric $g_{\mu\nu}$:

$$K^{\rho}_{\ \mu;\rho} = K^{\rho}_{\ \mu,\rho} + \left\{ \begin{matrix} \rho \\ \sigma\ \rho \end{matrix} \right\} K^{\sigma}_{\ \mu} - \left\{ \begin{matrix} \sigma \\ \mu\ \rho \end{matrix} \right\} K^{\rho}_{\ \sigma}. \tag{7.56}$$

The proof of these results is straightforward and is left as an exercise.

8. **The Schwarzschild solution.** In this chapter we shall study a rigorous solution of Einstein's field equations, the gravitational field of a static spherically symmetric mass distribution. This will provide a simple model of a star consisting of a sphere S, with matter inside and with vacuum outside (Figure 28).

A gravitational field is called *stationary* if space-time admits a one-parameter group of motions with a time-like Killing vector field. We can then choose coordinates so that

$$g_{mn,4} = 0. \tag{8.1}$$

A *static* field is a special case of a stationary field where the time-

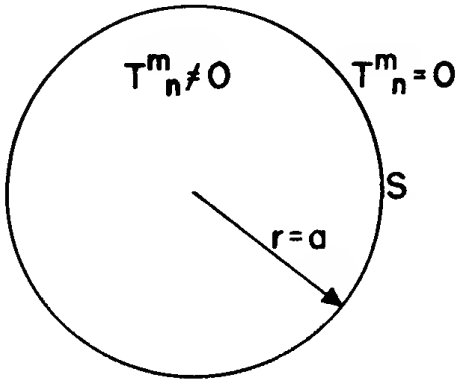


FIGURE 28

like Killing vectors are orthogonal to a family of hypersurfaces. Coordinates can then be chosen so that

$$g_{mn,4} = 0, \qquad g_{\mu 4} = 0. \tag{8.2}$$

A static field is called *spherically symmetric* if, in these coordinates, the 3-spaces $x^4 =$ constant have spherical symmetry. Such a field can be studied by analysing all the Killing vector fields which belong to the various symmetries.

We shall simplify the work by appealing to our geometrical intuition. We introduce a radial variable $\bar{x}^1 = \bar{r}$ and the spherical polar angles $x^2 = \theta$, $x^3 = \phi$ in the 3-spaces $x^4 = t =$ constant. Then the most general static and spherically symmetric space-time is given by the metric

$$ds^2 = -e^{\bar{\alpha}} d\bar{r}^2 - e^{\bar{\beta}} \bar{r}^2 (d\theta^2 + \sin^2\theta \, d\phi^2) + e^{\bar{\gamma}} dt^2, \tag{8.3}$$

where $\bar{\alpha}$, $\bar{\beta}$ and $\bar{\gamma}$ are functions of $\bar{r}$ alone.

We introduce a new radial variable $x^1 = r$ by the transformation

$$r^2 = e^{\bar{\beta}} \bar{r}^2, \tag{8.4}$$

which does not affect the other three coordinates. The metric (8.3) then simplifies to

$$ds^2 = -e^{\alpha} dr^2 - r^2 (d\theta^2 + \sin^2\theta \, d\phi^2) + e^{\gamma} dt^2, \tag{8.5}$$

where α and γ are functions of r. The new variable has a geometrical interpretation: the 2-sphere $t =$ constant, $r =$ constant has a surface area given by the Euclidean formula $4\pi r^2$.

Under the transformation (8.4), $\gamma(r)$ is obtained from the undifferentiated functions $\bar{\gamma}(\bar{r})$ and $\bar{\beta}(\bar{r})$, but $\alpha(r)$ involves $\bar{\alpha}(\bar{r})$, $\bar{\beta}(\bar{r})$ and the first derivative $d\bar{\beta}/d\bar{r}$. When we have a surface of discontinuity such as S, with $\bar{\alpha}$, $\bar{\beta}$, $\bar{\gamma}$ and their first derivatives continuous, then in the new variables α, γ and $d\gamma/dr$ will be continuous across S, but $d\alpha/dr$ can be discontinuous. We shall work with the metric (8.5) and not worry about the discontinuity of $d\alpha/dr$, because of the following fact (exercise): under a suitable transformation $r' = f(r)$ on one side of S, the full Lichnerowicz jump conditions ($g'_{mn} = [C]$, $g'_{mn,r} = [C]$) can be restored; in the new coordinates the metric will have the simple form (8.5) on one side of S, but the more complicated form (8.3) on the other side.

Our model star consists of a *perfect fluid* which is static and spherically symmetric. The energy momentum tensor is

$$T^m{}_n = (\rho + p)\, u^m u_n - p\delta^m{}_n, \tag{8.6}$$

where the proper density ρ and the proper pressure p are functions of r, and where

$$u^m = e^{-\gamma/2}(0,0,0,1), \qquad u_m = e^{\gamma/2}(0,0,0,1). \tag{8.7}$$

Thus the only nonzero components of $T^m{}_n$ are

$$T^1{}_1 = T^2{}_2 = T^3{}_3 = -p, \qquad T^4{}_4 = \rho. \tag{8.8}$$

Any tensor built up from the metric tensor and the stress energy tensor must share their symmetries. Since g_{mn} and T_{mn} are invariant under the transformation $t' = -t$, with the other coordinates unchanged, it follows that the components of such a tensor vanish which contain an odd number of suffixes 4. Similarly, the components vanish which contain an odd number of suffixes 2 or an odd number of suffixes 3. For a symmetric tensor $Q_{mn} = Q_{nm}$, the spherical symmetry also implies $Q_{33} = \sin^2\theta\, Q_{22}$. The only nonzero components of $Q^m{}_n$ are:

$$Q^1{}_1, \qquad Q^2{}_2 = Q^3{}_3, \qquad Q^4{}_4. \tag{8.9}$$

Consider the conservation equations

$$Q^m{}_{n;m} = 0 \tag{8.10}$$

for such a tensor Q_{mn}. This vector equation is identically satisfied for $n = 2, 3, 4$, and the only independent equation is

$$Q^m{}_{1;m} = Q^1{}_{1,1} + \left(\left\{ \begin{matrix} m \\ 1 \; m \end{matrix} \right\} - \left\{ \begin{matrix} 1 \\ 1 \; 1 \end{matrix} \right\} \right) Q^1{}_1 - \left\{ \begin{matrix} 4 \\ 1 \; 4 \end{matrix} \right\} Q^4{}_4 - \left(\left\{ \begin{matrix} 2 \\ 1 \; 2 \end{matrix} \right\} + \left\{ \begin{matrix} 3 \\ 1 \; 3 \end{matrix} \right\} \right) Q^2{}_2 = 0, \tag{8.11}$$

where Equations (8.9) have been used. As we shall see shortly, the coefficient of $Q^2{}_2$ in Equation (8.11) is not zero.

We shall now select a complete set of independent gravitational field equations. We first choose the independent component

$$T^m{}_{1;m} = 0 \tag{8.12}$$

of the conservation laws $T^m{}_{n;m} = 0$. If we put $Q^m{}_n = G^m{}_n + 8\pi T^m{}_n$, we obtain $Q^m{}_{n;m} = 0$ since $G^m{}_{n;m} = 0$ identically. Then, by Equation (8.11), the field equation $Q^2{}_2 = 0$ is a consequence of $Q^1{}_1 = Q^4{}_4 = 0$. Thus, in addition to Equation (8.12) we only require the two field equations

$$G^1{}_1 = -8\pi T^1{}_1, \tag{8.13}$$

$$G^4{}_4 = -8\pi T^4{}_4. \tag{8.14}$$

In order to obtain explicit expressions for the field equations, we first compute the Christoffel symbols for the metric (8.5). The easiest way is to write out the Euler-Lagrange equations

$$\frac{\partial L}{\partial x^m} - \frac{d}{du}\left(\frac{\partial L}{\partial \dot{x}^m} \right) = 0 \tag{8.15}$$

of the variational principle

$$\delta \int L \, du = 0, \qquad L = g_{mn} \dot{x}^m \dot{x}^n, \qquad \dot{x}^m = dx^m/du, \tag{8.16}$$

for a geodesic with affine parameter u, and to compare these to the equations

$$\frac{D\dot{x}^m}{du} = \ddot{x}^m + \left\{ \begin{matrix} m \\ r \; s \end{matrix} \right\} \dot{x}^r \dot{x}^s = 0. \tag{8.17}$$

The nonzero Christoffel symbols can then be read off. They are

$$\begin{Bmatrix} 1 \\ 1\ 1 \end{Bmatrix} = \frac{1}{2}\alpha', \quad \begin{Bmatrix} 1 \\ 2\ 2 \end{Bmatrix} = -re^{-\alpha}, \quad \begin{Bmatrix} 1 \\ 3\ 3 \end{Bmatrix} = -r\sin^2\theta e^{-\alpha},$$

$$\begin{Bmatrix} 1 \\ 4\ 4 \end{Bmatrix} = \frac{1}{2}e^{\gamma-\alpha}\gamma', \quad \begin{Bmatrix} 2 \\ 1\ 2 \end{Bmatrix} = \begin{Bmatrix} 3 \\ 1\ 3 \end{Bmatrix} = \frac{1}{r}, \tag{8.18}$$

$$\begin{Bmatrix} 2 \\ 3\ 3 \end{Bmatrix} = -\sin\theta\cos\theta, \quad \begin{Bmatrix} 3 \\ 2\ 3 \end{Bmatrix} = \cot\theta, \quad \begin{Bmatrix} 4 \\ 1\ 4 \end{Bmatrix} = \frac{1}{2}\gamma',$$

where a prime denotes differentiation with respect to r.

Equation (8.12) now becomes

$$p' + \frac{1}{2}\gamma'(p+\rho) = 0. \tag{8.19}$$

A straightforward calculation gives $G^1{}_1$ and $G^4{}_4$. Equation (8.14) becomes

$$(re^{-\alpha})' - 1 = -8\pi r^2\rho, \tag{8.20}$$

and the difference between Equations (8.13) and (8.14)

$$e^{-\alpha}(\alpha' + \gamma') = 8\pi r(p+\rho). \tag{8.21}$$

If an equation of state is added, a relation between the pressure p and the density ρ, then this gives a system of four equations for the four unknown functions α, γ, p and ρ of the variable r.

Exercise. Show that the field equation (8.13) reads

$$-\frac{1}{r^2} + e^{-\alpha}\left(\frac{1}{r^2} + \frac{1}{r}\gamma'\right) = 8\pi p. \tag{8.22}$$

The Exterior Solution.[21] This is the vacuum solution which holds in the exterior of our model star. The field equations are (8.20) and (8.21) with $p = \rho = 0$. We also require that at spatial infinity, where the gravitational field becomes weak, the metric (8.5) approach the Minkowski metric of special relativity theory for spherical polar coordinates. This gives the boundary conditions

$$\alpha \to 0, \quad \gamma \to 0, \quad \text{as } r \to \infty. \tag{8.23}$$

The first field equation (8.20) gives

[21] For a vacuum field, spherical symmetry implies that the field is static; this is known as Birkhoff's theorem. It is proved in J. L. Synge, *Relativity: The general theory*, North-Holland Publishing Co., Amsterdam, 1960.

$$e^{-\alpha} = 1 - \frac{2m}{r}, \tag{8.24}$$

where m is a constant of integration. This automatically satisfies Equation (8.23). In Chapter 4 we saw that $g_{44} = 1 + 2V$ in the Newtonian approximation, V being the Newtonian gravitational potential. For a sphere of mass m, $V = -m/r$. We therefore call m in Equation (8.24) the *mass* of the star. For large values of r, where the gravitational field is weak, m behaves like the Newtonian mass. We shall verify this again in the following chapter when we study planetary motion.

The second field equation (8.21) gives $\alpha' + \gamma' = 0$, and therefore

$$\gamma = -\alpha, \tag{8.25}$$

the constant of integration being zero by Equation (8.23).

Schwarzschild's vacuum metric is now completely determined:

$$ds^2 = -\frac{dr^2}{1 - 2m/r} - r^2(d\theta^2 + \sin^2\theta\, d\phi^2) + \left(1 - \frac{2m}{r}\right) dt^2. \tag{8.26}$$

The metric becomes singular on the *Schwarzschild sphere* of radius

$$r_s = 2m. \tag{8.27}$$

This is indeed a surface with peculiar physical properties. However, mathematically it is *not* a singularity of space-time but merely a singularity of the coordinate system. This will be shown in a later section.

For an ordinary body, a star, a planet or a nucleon, the Schwarzschild radius is always much smaller than the size of the body. Therefore the vacuum solution (8.26) does not hold at the Schwarzschild radius $2m$, and there is no Schwarzschild sphere. In our units ($G = c = 1$), mass may be measured in units of length. Here are some numerical examples: $m(\text{Sun}) = 1.5\,\text{km}$, $m(\text{Earth}) = 0.44\,\text{cm}$, $m(\text{Neutron}) = 1.2 \times 10^{-54}\,\text{cm}$.

Exercise. (1) Show that, under the transformation $r = (1 + m/2\bar{r})^2\bar{r}$, the Schwarzschild metric (8.26) becomes

$$ds^2 = -\left(1 + \frac{m}{2\bar{r}}\right)^4 [d\bar{r}^2 + \bar{r}^2(d\theta^2 + \sin^2\theta\, d\phi^2)] + \frac{(1 - m/2\bar{r})^2}{(1 + m/2\bar{r})^2} dt^2. \tag{8.28}$$

The new coordinates are called *isotropic coordinates.*

(2) Show that, under the transformation $t = \bar{t} + 2m \log(r - 2m)$, the Schwarzschild metric (8.26) becomes

$$ds^2 = -\left(1 + \frac{2m}{r}\right) dr^2 - r^2(d\theta^2 + \sin^2\theta\, d\phi^2) + \left(1 - \frac{2m}{r}\right) d\bar{t}^2 + \frac{4m}{r} dr\, d\bar{t}. \tag{8.29}$$

The Interior Solution. We shall obtain the gravitational field in the interior of our model star for the simple case where the fluid is homogeneous:

$$\rho = \text{constant}. \tag{8.30}$$

We shall also assume that the Schwarzschild sphere of the exterior solution is buried well within the star, i.e., $a \gg r_s = 2m$.

First let us state the junction conditions which must be satisfied on S (Figure 28) where $r = a$. Clearly

$$\alpha = [C], \quad \gamma = [C]. \tag{8.31}$$

The field equation (8.22) then shows that $\gamma' = [C]$ is equivalent to $p = [C]$ and, since p vanishes in the exterior, to

$$p = 0 \tag{8.32}$$

on S for the interior solution. The junction conditions (7.13) are immediately seen to reduce to just this equation.

Finally, Equation (8.20) shows that since ρ is discontinuous across S, α' is discontinuous too. As remarked earlier, the full junction conditions $g_{mn} = [C]$, $g_{mn,r} = [C]$ can be satisfied, but only by performing a coordinate transformation after which the metric has the more complicated form (8.3) on one side of S. We shall retain the metric (8.5) and give up the continuity of α'.

We shall now solve the field equations. Equation (8.20) gives

$$re^{-\alpha} = r - \frac{8\pi}{3} \rho r^3 + B. \tag{8.33}$$

The constant of integration B is zero, since otherwise e^{α} would vanish at the origin $r = 0$ and the metric would become singular. Thus

$$e^{\alpha} = \left(1 - \frac{8\pi}{3}\rho r^2\right)^{-1}. \tag{8.34}$$

Comparison with the exterior solution (8.26) and the continuity of e^{α} at $r = a$ give

$$m = \frac{4\pi}{3}\rho a^3. \tag{8.35}$$

Equation (8.19) gives

$$e^{\gamma} = A^2(p + \rho)^{-2}. \tag{8.36}$$

The boundary conditions $e^{\gamma} = [C]$ and $p = 0$ for $r = a$, determine the constant of integration A:

$$\frac{A}{\rho} = \left(1 - \frac{8\pi}{3}\rho a^2\right)^{1/2}. \tag{8.37}$$

Substituting from Equations (8.34) and (8.36) into the last field equation (8.21), we obtain a linear differential equation for $e^{\gamma/2}$. Its solution is

$$e^{\gamma/2} = \frac{3}{2}\frac{A}{\rho} + C\left(1 - \frac{8\pi}{3}\rho r^2\right)^{1/2}. \tag{8.38}$$

The constant of integration C is determined from the continuity of e^{γ} at $r = a$; it is

$$C = -1/2. \tag{8.39}$$

This completes the solution. Written out in full, it is

$$\begin{aligned} ds^2 = &-\left(1 - \frac{8\pi}{3}\rho r^2\right)^{-1} dr^2 - r^2(d\theta^2 + \sin^2\theta\, d\phi^2) \\ &+ \left[\frac{3}{2}\left(1 - \frac{8\pi}{3}\rho a^2\right)^{1/2} - \frac{1}{2}\left(1 - \frac{8\pi}{3}\rho r^2\right)^{1/2}\right]^2 dt^2, \quad r \leqq a, \end{aligned}$$

$$\begin{aligned} ds^2 = &-\left(1 - \frac{2m}{r}\right)^{-1} dr^2 - r^2(d\theta^2 + \sin^2\theta\, d\phi^2) \\ &+ \left(1 - \frac{2m}{r}\right) dt^2, \quad r \geqq a, \end{aligned} \tag{8.40}$$

$$m = \frac{4\pi}{3}\rho a^3.$$

The pressure distribution in the star is given by

$$\frac{p}{\rho} = \frac{\left(1 - \frac{8\pi}{3}\rho r^2\right)^{1/2} - \left(1 - \frac{8\pi}{3}\rho a^2\right)^{1/2}}{3\left(1 - \frac{8\pi}{3}\rho a^2\right)^{1/2} - \left(1 - \frac{8\pi}{3}\rho r^2\right)^{1/2}}, \quad r \leqq a. \tag{8.41}$$

If the pressure is not to become infinite for real values of r we must have (exercise):

$$a > \frac{9}{4}m, \tag{8.42}$$

a slightly stronger condition than $a > 2m$.

The Kruskal Extension. Kruskal[22] has given an analytic extension of the spherically symmetric vacuum solution across the Schwarzschild sphere. This is of intrinsic theoretical interest, but it may also have applications if large masses collapse gravitationally in peculiar extragalactic objects such as quasi-stellar sources.[23]

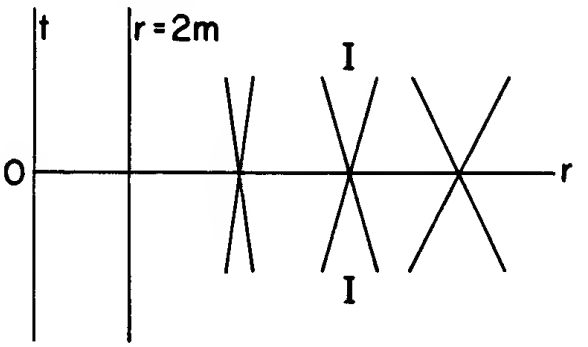


FIGURE 29

Let us concentrate on the rt-plane of the exterior Schwarzschild metric (8.26). The light cone collapses into the t-direction as r approaches the Schwarzschild radius $2m$ (Figure 29). This is avoided by going to coordinates u, v in which radial light rays have the slopes ± 1 everywhere. Such coordinates are given by

[22] M. D. Kruskal, Phys. Rev. **119** (1960), 1743.

[23] F. Hoyle and W. A. Fowler, M. N. Royal Astron. Soc. **125** (1963), 169; Nature **197** (1963), 533. See also *Quasi-stellar sources and gravitational collapse*, Proceedings of the First Texas Symposium on Relativistic Astrophysics, Univ. Chicago Press, Chicago, Ill., 1965.

$$u = \left(\frac{r}{2m} - 1\right)^{1/2} e^{r/4m} \cosh \frac{t}{4m},$$
$$v = \left(\frac{r}{2m} - 1\right)^{1/2} e^{r/4m} \sinh \frac{t}{4m}. \tag{8.43}$$

The Schwarzschild metric (8.26) becomes

$$ds^2 = \frac{32m^3}{r} e^{-r/2m}(-du^2 + dv^2) - r^2(d\theta^2 + \sin^2\theta\, d\phi^2), \tag{8.44}$$

where r is a function of $u^2 - v^2$ given by

$$\left(\frac{r}{2m} - 1\right) e^{r/2m} = u^2 - v^2. \tag{8.45}$$

From Equation (8.43) we also have

$$\tanh \frac{t}{4m} = \frac{v}{u}. \tag{8.46}$$

At the Schwarzschild radius $r = 2m$, $u = \pm v$ and the metric (8.44) is well behaved. The new coordinates u, v, θ, ϕ thus cover a larger domain of space-time than did the coordinates t, r, θ, ϕ. Figure 30 shows the uv-plane. The quadrant I corresponds to the region $r > 2m$ in the rt-plane (Figure 29). The quadrants II, III and IV are new regions of space-time in which the gravitational vacuum solutions apply. If we define r and t in these regions by Equations (8.45) and (8.46), then throughout the uv-plane, radial lines through the origin correspond to $t =$ constant and hyperbolas $u^2 - v^2 =$ constant to $r =$ constant. The two branches of the hyperbola

$$r = 0, \qquad v^2 - u^2 = 1, \tag{8.47}$$

which lie in the quadrants II and III are singularities of space-time. That they are intrinsic geometrical singularities, and not just singularities of the coordinates, can be demonstrated by showing that invariants constructed from the curvature tensor become infinite there. Thus the Kruskal extension is a maximal analytic extension of the exterior Schwarzschild solution.

In studying Figure 30 it must be kept in mind that each point represents a sphere. For example, the origin $u = v = 0$ is a sphere of "radius" $r = 2m$ in the sense that it has surface area $4\pi r^2$. We

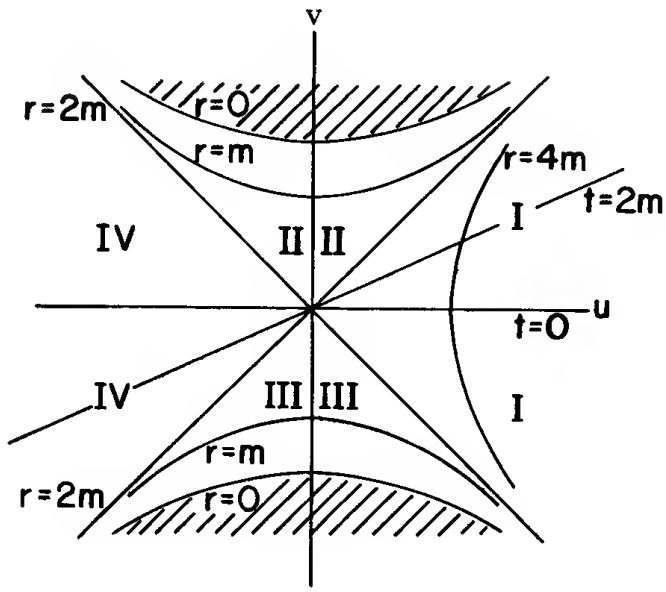


FIGURE 30

illustrate this with the hypersurface $v = 0$. Suppressing one dimension of space, this is drawn in Figure 31 as a surface in Euclidean 3-space; the spheres $r =$ constant are shown as circles.

The following properties of the extended space-time are easily seen to hold.

1. In the regions II and III (Figure 30), space-time is *not* static, or even stationary. There are, of course, Killing vectors pointing along the curves r, θ, $\phi =$ const, but these are now in space-like directions.

2. The Schwarzschild sphere (8.47) is a null hypersurface which separates region I (or IV) from regions II and III.

3. Particles and light signals can penetrate the Schwarzschild sphere from region I to II, but *not* from region II to I.

4. Particles and light signals can penetrate the Schwarzschild sphere from region III to I, but *not* from I to III.

5. Regions I and IV are causally disconnected. Neither particles nor signals can move from one of them to the other.

6. A particle can move from region I to region II in *finite* proper time. Examples of such a motion are radial free fall (radial geodesic) or the time-like world line $u =$ a positive constant. The particle must finally hit the true singularity $r = 0$. Consider an observer

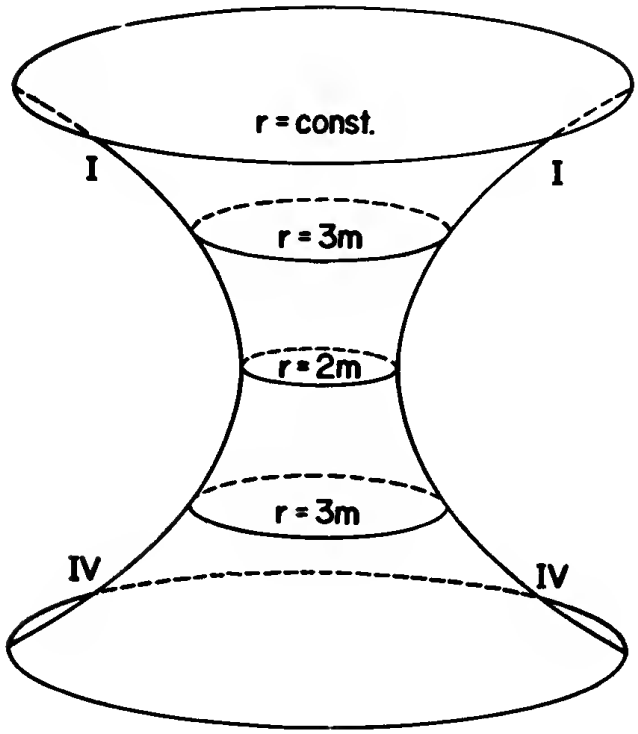


FIGURE 31

in region I, either at rest in the static field (r, θ, ϕ = constants) or in a planetary orbit. Such an observer *never* sees the particle cross into region II. An infinite amount of the observer's proper time elapses while he sees the particle approach the Schwarzschild sphere.

9. **The classical tests.** In this chapter we shall study the three classical tests of general relativity theory. The gravitational field is in each case the exterior Schwarzschild field for large values of r ($2m/r \ll 1$). However, in order to see which terms enter each effect, and also to make possible the comparison with gravitational theories other than Einstein's, we shall work with a general static, spherically symmetric metric. We take

$$\begin{aligned} ds^2 = &- (1+G)\,dr^2 - (1+H)\,r^2(d\theta^2 + \sin^2\theta\, d\phi^2) \\ &+ (1+F)\,dt^2 + 2J\,dr\,dt, \end{aligned} \tag{9.1}$$

where F, G, H and J are functions of the small quantity $2m/r$, which have the power series expansions

$$
\begin{aligned}
F &= -\left(\frac{2m}{r}\right) + \alpha_2\left(\frac{2m}{r}\right)^2 + \cdots, \\
G &= \beta_1\left(\frac{2m}{r}\right) + \beta_2\left(\frac{2m}{r}\right)^2 + \cdots, \\
H &= \gamma_1\left(\frac{2m}{r}\right) + \gamma_2\left(\frac{2m}{r}\right)^2 + \cdots, \\
J &= \delta_1\left(\frac{2m}{r}\right) + \delta_2\left(\frac{2m}{r}\right)^2 + \cdots.
\end{aligned} \tag{9.2}
$$

There is no loss of generality in the choice of the coefficient $\alpha_1 = -1$ in the expansion of F. This choice merely ensures that m is the Newtonian mass of the attracting body in our units, and not some other constant proportional to the mass. An exception is the case $\alpha_1 = 0$ which is not considered here.

The Schwarzschild metric (8.26) corresponds to the parameter values

$$
\begin{aligned}
\alpha_2 &= \alpha_3 = \cdots = 0, \\
\beta_1 &= \beta_2 = \cdots = 1, \\
\gamma_1 &= \gamma_2 = \cdots = 0, \\
\delta_1 &= \delta_2 = \cdots = 0.
\end{aligned} \tag{9.3}
$$

Exercise. Show that the metric (9.1), (9.2) is static, and not merely stationary.

The Red Shift. We first give the general theory of the frequency shift of spectral lines. Let Γ_0 be the world line of a source which emits light of frequency ν_0, and let Γ be the world line of an observer receiving the light at frequency ν (Figure 32).

From two neighboring events A_0 and B_0 on Γ_0 draw the two null geodesics into the future which intersect Γ in the two neighboring events A and B. If we ignore pathological situations, the construction will be unique and the ratio ds_0/ds of the proper times will be completely determined by the two world lines Γ_0 and Γ, and by the event A_0 (or, equivalently, by A). Null geodesics are curves of constant phase for light which propagates along them. Therefore the phase difference between A_0 and B_0 will be the same as between A and B. The period of the light will be proportional to the time elapsed between the two events of fixed phase difference,

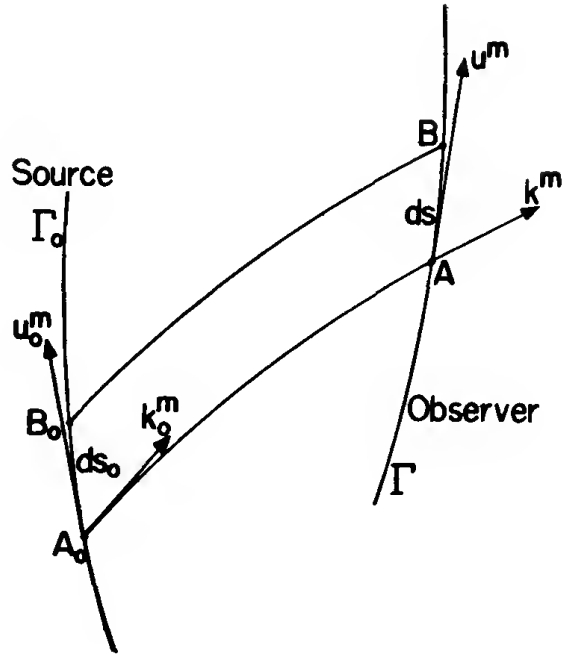


FIGURE 32

the frequency will be inversely proportional to the elapsed time. For the source the elapsed time is the proper time ds_0, for the observer the proper time ds. We therefore have the general red shift formula

$$\frac{\nu}{\nu_0} = \frac{ds_0}{ds}. \tag{9.4}$$

We may call ν_0 the *proper frequency*. For an atom moving with the source and emitting a particular spectral line, ν_0 is the frequency found in tables of physical constants. It is the same frequency which the observer would measure in his own laboratory for the same atom moving with the observer. The *observed frequency* ν is the frequency which the observer sees when he looks through a telescope at the source.

The frequency shift formula (9.4) gives the total shift which involves the motion of source and observer as well as the curvature of space-time. In the general case there is no natural way of separating this into a velocity Doppler shift due to motion and a gravitation shift due to curvature.

Exercise. A wave vector k_0^m can be defined at A_0, so that k_0^m is

tangent to the null geodesic A_0A and $\nu_0 = k_{0m}u_0^m$ (Figure 32). Show that, in the proper frame of the source, this agrees with the definition (2.7) of a wave vector in special relativity theory. Show that the formula

$$\frac{\nu}{\nu_0} = \frac{k_m u^m}{k_{0m}u_0^m} \tag{9.5}$$

gives the same red shift as Equation (9.4), if k^m at A is obtained by parallel propagation of k_0^m at A_0 along the null geodesic A_0A.

We now consider the very special case where the gravitational field is stationary, the coordinates satisfy $g_{mn,4} = 0$, and both source and observer are at rest, i.e., x_0^μ = constants along Γ_0, and x^μ = constants along Γ. Then the null geodesic B_0B is obtained from A_0A by a constant infinitesimal displacement in the x^4- or t-direction. It follows that $dt_0 \equiv dx^4(A_0B_0) = dt \equiv dx^4(AB)$. The red shift formula (9.4) simplifies to

$$\frac{\nu}{\nu_0} = \left(\frac{g_{44}(A_0)}{g_{44}(A)}\right)^{1/2}. \tag{9.6}$$

For the metric (9.1) this gives to the lowest order in the small quantity $2m/r$:

$$\frac{\nu_0 - \nu}{\nu_0} = \frac{m}{r_0} - \frac{m}{r}. \tag{9.7}$$

This, in slightly different notation, is the same result which in Equation (3.11) was obtained from the equivalence principle. It is therefore not surprising that the gravitational red shift involves only the Newtonian term in g_{44}.

For light coming from near the sun's surface, the gravitational red shift is 2.12×10^{-6}. J. Brault[24] has measured this with an accuracy of about 5%, using the sodium D line which comes from below the sun's turbulent and high temperature upper atmosphere.

On the earth's surface the red shift is 1.09×10^{-16} for a one meter difference in height between source and observer. Pound[25] and his coworkers have measured this with an accuracy of about 1%, using the Mössbauer effect.

---

[24] R. H. Dicke, *The theoretical significance of experimental relativity*, Gordon and Breach, New York, 1964.

[25] Op. Cit.

Planetary Motion; Perihelion Rotation. To a very good approximation, a planet is a test particle without spin and moves along a geodesic.

The geodesic equations for our metric (9.1) are obtained from the Euler-Lagrange equations (8.15) of the variational principle (8.16). The equation for $x^2 = \theta$ is satisfied by $\theta = \pi/2$. This merely tells us that the particle's orbit is plane and that, having chosen the polar coordinates so that the particle starts moving in the equatorial plane ($\theta = \pi/2$, $\dot{\theta} = 0$ initially), the particle will remain in the equatorial plane. We therefore simply put $\theta = \pi/2$, without loss of generality. Since g_{mn} does not contain the variables t and ϕ, the equations for $x^4 = t$ and $x^3 = \phi$ immediately yield the first integrals

$$(1 + F)\,\dot{t} + J\dot{r} = C, \tag{9.8}$$

$$(1 + H)\,r^2\dot{\phi} = h, \tag{9.9}$$

the dot denoting differentiation d/du with respect to an affine parameter u. Finally, the equation for $x^1 = r$ can be replaced by the first integral $g_{mn}\dot{x}^m\dot{x}^n = \epsilon$:

$$-(1 + G)\,\dot{r}^2 - (1 + H)\,r^2\dot{\phi}^2 + (1 + F)\,\dot{t}^2 + 2J\dot{r}\dot{t} = \epsilon. \tag{9.10}$$

In these equations, C, h and ϵ are constants of the motion.

Introducing the reciprocal radius

$$\rho = 1/r, \tag{9.11}$$

and eliminating the variables t and u, we obtain the equation for the orbit:

$$\begin{aligned}[(1 + G)(1 + F) + J^2]\left(\frac{d\rho}{d\phi}\right)^2 \\ = -(1 + H)(1 + F)\,\rho^2 + (1 + H)^2\frac{2e}{h^2} - \frac{\epsilon}{h^2}(1 + H)^2F,\end{aligned} \tag{9.12}$$

where $2e = C^2 - \epsilon$.

Up to this point we have kept the constant ϵ general because, for $\epsilon = 0$, the above equations are those of a null geodesic and will be used in the following section to analyse the bending of light. Now, however, we shall limit ourselves to the motion of a particle of positive mass. Then $\epsilon > 0$ and we can, without loss of generality,

put

$$\epsilon = 1. \tag{9.13}$$

This means that $u = s$, instead of the parameter u being merely proportional to s.

For slow motion, when

$$v^2 = \left(\frac{dr}{dt}\right)^2 + r^2\left(\frac{d\phi}{dt}\right)^2 \tag{9.14}$$

and $2m\rho$ are small compared to unity, and of the same order, we obtain in first approximation the following: Equation (9.10) gives

$$\dot{t}^2 = 1 + v^2 + \frac{2m}{r}; \tag{9.15}$$

Equations (9.8), (9.9) and (9.12) give

$$e = \frac{1}{2}v^2 - \frac{m}{r}, \tag{9.16}$$

$$h = r^2\frac{d\phi}{dt}, \tag{9.17}$$

$$\left(\frac{d\rho}{d\phi}\right)^2 = -\rho^2 + \frac{2e}{h^2} + \frac{2m}{h^2}\rho. \tag{9.18}$$

The last three equations are exactly those of the one-body problem in Newtonian gravitational theory, e and h being respectively the total energy and the angular momentum per unit mass of the moving particle. This confirms the result of Chapter 4, that general relativity theory reduces to Newtonian theory in lowest approximation.

Since the right-hand side of Equation (9.18) is a quadratic polynomial in ρ, it can be written

$$\left(\frac{d\rho}{d\phi}\right)^2 = (\rho - \rho_1)(\rho_2 - \rho), \tag{9.19}$$

and since it must be nonnegative, $\rho_1 \leqq \rho \leqq \rho_2$ for elliptic orbits; ρ_1 and ρ_2 are the reciprocals of the apsidal distances.

In second approximation, Equation (9.12) gives

$$\left(\frac{d\rho}{d\phi}\right)^2 = -\rho^2 + \frac{2e}{h^2} + \frac{2m}{h^2}\rho + 2(\beta_1 - \gamma_1)m\rho^3$$

$$+ \frac{4}{h^2}(1 + 2\gamma_1 - \beta_1 - \alpha_2)m^2\rho^2 + \frac{4e}{h^2}(1 + 2\gamma_1 - \beta_1)m\rho \tag{9.20}$$

$$= 2(\beta_1 - \gamma_1)m(\rho - \rho_1)(\rho_2 - \rho)(\rho_3 - \rho),$$

where ρ_1 and ρ_2 are close to the Newtonian values of Equation (9.19) and ρ_3 is much larger. For orbits which are close to the elliptic orbits of the Newtonian first approximation, $\rho_1 \leqq \rho \leqq \rho_2 \ll \rho_3$, and ρ_1 and ρ_2 are again the reciprocals of the apsidal radial coordinates (Figure 33). The apsidal angle, the angle between two

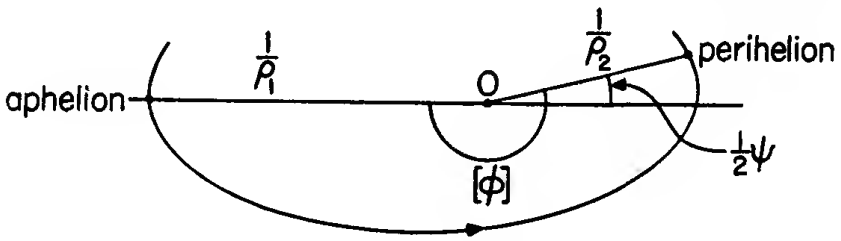


FIGURE 33

successive apsidal radii, is obtained by integrating Equation (9.20):

$$[\Phi] = \int_{\rho_1}^{\rho_2} \frac{d\rho}{[(\rho - \rho_1)(\rho_2 - \rho)]^{1/2}} [2(\beta_1 - \gamma_1)m(\rho_3 - \rho)]^{-1/2}. \tag{9.21}$$

Equating the coefficients of ρ^2 in the two right-hand sides of Equation (9.20), we have

$$-2(\beta_1 - \gamma_1)m(\rho_1 + \rho_2 + \rho_3)$$

$$= -1 + 4\frac{m^2}{h^2}(1 + 2\gamma_1 - \beta_1 - \alpha_2), \tag{9.22}$$

or

$$2(\beta_1 - \gamma_1)m(\rho_3 - \rho)$$

$$= 1 - (\beta_1 - \gamma_1)m(2\rho - \rho_1 - \rho_2) - 3(\beta_1 - \gamma_1)m(\rho_1 + \rho_2) \tag{9.23}$$

$$- \frac{4m^2}{h^2}(1 + 2\gamma_1 - \beta_1 - \alpha_2).$$

Substituting into Equation (9.21), and keeping only terms of the order required, we have

$$[\Phi] = \int_{\rho_1}^{\rho_2} \frac{d\rho}{[(\rho-\rho_1)(\rho_2-\rho)]^{1/2}} \Big[1 + \frac{1}{2}(\beta_1 - \gamma_1) m(2\rho - \rho_1 - \rho_2) + \frac{3}{2}(\beta_1 - \gamma_1) m(\rho_1 + \rho_2) + \frac{2m^2}{h^2}(1 + 2\gamma_1 - \beta_1 - \alpha_2) \Big]$$

$$= \pi + \frac{3}{2}\pi(\beta_1 - \gamma_1) m(\rho_1 + \rho_2) + \frac{2\pi m^2}{h^2}(1 + 2\gamma_1 - \beta_1 - \alpha_2). \tag{9.24}$$

Since $[\Phi]$ is the apsidal angle for half a revolution in the Newtonian orbit, the angle of rotation of the perihelion per revolution is given by $\psi = 2[\Phi] - 2\pi$. Also, since the last two terms in Equation (9.24) are small, we can use the Newtonian values

$$\frac{2m}{h^2} = \rho_1 + \rho_2 = \frac{2}{(1-e^2)a}, \tag{9.25}$$

where a is the semimajor axis and e the eccentricity of the elliptic orbit of the Newtonian approximation. The final result is that the perihelion of the particle rotates in the same sense as the orbital motion through an angle

$$\psi = \frac{2\pi m}{(1-e^2)a}(2 + \gamma_1 + \beta_1 - 2\alpha_2) \tag{9.26}$$

per revolution. We see that this expression involves not only the first order terms in g_{44}, g_{11}, g_{22}, but also the second order term in g_{44}.

For the Schwarzschild metric with parameter values (9.3) this becomes

$$\psi = \frac{6\pi m}{(1-e^2)a}. \tag{9.27}$$

From the physical and geometrical interpretation of this effect, it is clear that the expression for ψ must be independent of the particular form used for the Schwarzschild metric.

Exercise. Check that Equation (9.27) is obtained for the metric (8.28) or (8.29).

Duncombe[26] gives the following results for the perihelion rotations of Mercury, Venus and the Earth in seconds of arc per century:

[26] R. L. Duncombe, Astronom. J. **61** (1956), 174.

	Mercury	*Venus*	*Earth*
Observed	$43''.11 \pm .45$	$8''.4 \pm 4.8$	$5''.0 \pm 1.2$
Theoretical	$43''.03$	$8''.6$	$3''.8$

Dicke[27] points out that other factors, such as a solar quadrupole moment, might reduce the agreement between observation and theory.

The Bending of Light. A light pulse moves along a null geodesic. Its orbit is given by Equation (9.12) with $\epsilon = 0$:

$$[(1+G)(1+F)+J^2]\left(\frac{d\rho}{d\phi}\right)^2 = -(1+H)(1+F)\rho^2+(1+H)^2\frac{C^2}{h^2}. \tag{9.28}$$

In the first approximation, this is

$$\left(\frac{d\rho}{d\phi}\right)^2 = -\rho^2+\frac{C^2}{h^2}. \tag{9.29}$$

For a suitable choice of origin of the polar angle ϕ, the solution is the straight line

$$\rho = \frac{\cos\phi}{R}, \tag{9.30}$$

where $R^2 = h^2/C^2$.

In the second approximation, Equation (9.28) is

$$\left(\frac{d\rho}{d\phi}\right)^2 = -\rho^2+\frac{1}{R^2}+2(\beta_1-\gamma_1)m\rho^3+\frac{2}{R^2}(1+2\gamma_1-\beta_1)m\rho. \tag{9.31}$$

Differentiating with respect to ϕ, we obtain

$$\frac{d^2\rho}{d\phi^2}+\rho = 3(\beta_1-\gamma_1)m\rho^2+\frac{m}{R^2}(1+2\gamma_1-\beta_1). \tag{9.32}$$

In the small terms on the right-hand side of this equation we can substitute from the first approximation (9.30) and obtain

[27] R. H. Dicke, *The theoretical significance of experimental relativity*, Gordon & Breach, New York, 1964.

$$\frac{d^2\rho}{d\phi^2} + \rho = \frac{3m}{R^2}(\beta_1 - \gamma_1)\cos^2\phi + \frac{m}{R^2}(1 + 2\gamma_1 - \beta_1). \tag{9.33}$$

The solution of this equation, which in first approximation reduces to (9.30), is

$$\rho = \frac{\cos\phi}{R} + \frac{m}{R^2}(\beta_1 - \gamma_1)(\cos^2\phi + 2\sin^2\phi) + \frac{m}{R^2}(1 + 2\gamma_1 - \beta_1). \tag{9.34}$$

If we introduce Cartesian coordinates (more precisely, coordinates which become Cartesian at infinity) $x = r\cos\phi = (\cos\phi)/\rho$, $y = r\sin\phi = (\sin\phi)/\rho$, this becomes

$$\begin{aligned} R = x &+ \frac{m}{R}(\beta_1 - \gamma_1)\frac{x^2 + 2y^2}{(x^2 + y^2)^{1/2}} \\ &+ \frac{m}{R}(1 + 2\gamma_1 - \beta_1)(x^2 + y^2)^{1/2}. \end{aligned} \tag{9.35}$$

The asymptotes of this curve are obtained by taking y large in magnitude compared to x. They are given by

$$x = R \pm \frac{m}{R}(1 + \beta_1)\, y. \tag{9.36}$$

The orbit of the null geodesic is shown in Figure 34. The angle

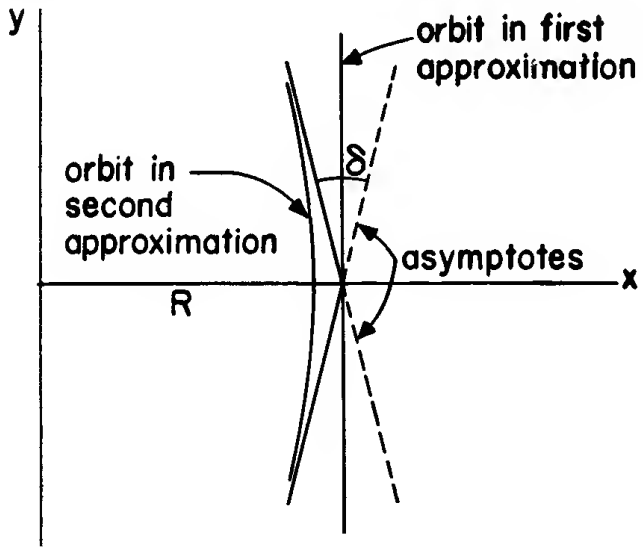


FIGURE 34

through which the orbit turns is the angle between the asymptotes:

$$\delta = \frac{2m}{R}(1 + \beta_1). \tag{9.37}$$

This effect involves the Newtonian term in g_{44} and the first order term in g_{11}.

For the Schwarzschild metric the angle through which light is bent is

$$\delta = 4m/R. \tag{9.38}$$

This effect predicts that stars seen near the surface of the sun should seem displaced through the angle δ. For light grazing the limb of the sun, $\delta = 1''.75$. The observations are difficult and, so far, have only been made during solar eclipses. They agree with the predicted effect with an uncertainty of some 10-20%.[28]

10. **The Petrov-Pirani classification.** The algebraic classification of the vacuum Riemann tensor has in recent years become a useful tool in the investigation of gravitational radiation and in the discovery and classification of exact solutions of Einstein's field equations. The most systematic treatment of this problem is by means of spinors.[29]

Spinors in Curved Space-Time. We introduce an orthonormal tetrad $e_\alpha{}^m$ at each point of space-time, $e_4{}^m$ being a time-like vector pointing into the future. Greek suffixes will label the different tetrad vectors, and will range and sum over 1, 2, 3, 4. Tetrad labels will be raised and lowered with the Minkowski metric $\eta_{\alpha\beta} = \eta^{\alpha\beta} = \mathrm{diag}(-1, -1, -1, 1)$, e.g., $e^{\alpha m} = \eta^{\alpha\beta} e_\beta{}^m$. Latin suffixes are tensor suffixes and are raised and lowered with the metric tensors g_{mn} and g^{mn}.

The orthonormality of the tetrad gives

$$e_\alpha{}^m e^\beta{}_m = \delta^\beta{}_\alpha \Leftrightarrow e_\alpha{}^m e^\alpha{}_n = \delta^m{}_n. \tag{10.1}$$

It follows (exercise) that $\det(e^\alpha{}_m) = \pm(-g)^{1/2}$. We assume that the orientation of the tetrad is chosen so that the plus sign holds:

[28] R. H. Dicke, *The theoretical significance of experimental relativity*, Gordon & Breach, New York, 1964.

[29] R. Penrose, Ann. Phys. **10** (1960), 171; F. A. E. Pirani, *Lectures on general relativity*, 1964 Brandeis Summer Institute in Theoretical Physics, Prentice-Hall, N. J., 1965.

$$\det(e^{\alpha}{}_{m}) = (-g)^{1/2}. \tag{10.2}$$

We shall limit ourselves to coordinate transformations with positive Jacobian, so that this relation persists.

The most general change from one such tetrad field to another is given by

$$e'^{\alpha m} = L^{\alpha}{}_{\beta} e^{\beta m}, \tag{10.3}$$

where $L^{\alpha}{}_{\beta}(x)$ is a function of position which at each point satisfies the conditions of a proper, orthochronous Lorentz transformation:

$$\begin{gathered}\eta_{\alpha\beta} L^{\alpha}{}_{\mu} L^{\beta}{}_{\nu} = \eta_{\mu\nu} \Leftrightarrow \eta^{\mu\nu} L^{\alpha}{}_{\mu} L^{\beta}{}_{\nu} = \eta^{\alpha\beta}, \ \det(L^{\alpha}{}_{\beta}) = 1, \\ L^{4}{}_{4} \geqq 1.\end{gathered} \tag{10.4}$$

A tensor $T^{m}{}_{n}{}^{\cdots}{}_{\cdots}$ can be characterized completely by its *tetrad components* $T^{\mu}{}_{\nu}{}^{\cdots}{}_{\cdots}$, defined by

$$T^{\mu}{}_{\nu}{}^{\cdots}{}_{\cdots} = e^{\mu}{}_{m} e_{\nu}{}^{n} \cdots T^{m}{}_{n}{}^{\cdots}{}_{\cdots} \Leftrightarrow T^{m}{}_{n}{}^{\cdots}{}_{\cdots} = e_{\mu}{}^{m} e^{\nu}{}_{n} \cdots T^{\mu}{}_{\nu}{}^{\cdots}{}_{\cdots}. \tag{10.5}$$

We shall underline a tetrad component with a numerical suffix to distinguish it from the corresponding tensor component; thus $T^{\underline{2}}$ is a component of T^{μ}, T^{2} a component of T^{m}.

In a curved space-time spinors are associated with tetrad components of tensors in exactly the same way in which they were associated with rectangular Cartesian tensor components in Minkowski space. This is equivalent to defining the association $T_{a\cdots}{}^{b\cdots} \leftrightarrow T_{AC'BD'\cdots}$ by

$$\begin{aligned} T_{a\cdots}{}^{b\cdots} &= \sigma_{a}{}^{AC'} \sigma^{bBD'} \cdots T_{AC'BD'\cdots} \\ &\Leftrightarrow T_{AC'BD'\cdots} = \sigma^{a}{}_{AC'} \sigma_{bBD'} \cdots T_{a\cdots}{}^{b\cdots} \end{aligned} \tag{10.6}$$

where

$$\sigma_{a}{}^{AB'} = e^{\alpha}{}_{a} \sigma_{\alpha}{}^{AB'}, \tag{10.7}$$

and where $\sigma_{\alpha}{}^{AB'}$ are the constant Hermitian matrices given in Equation (1.25):

$$\begin{aligned} \sigma_{\underline{1}}{}^{AB'} &= \frac{1}{2^{1/2}} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_{\underline{2}}{}^{AB'} = \frac{1}{2^{1/2}} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \\ \sigma_{\underline{3}}{}^{AB'} &= \frac{1}{2^{1/2}} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \sigma_{\underline{4}}{}^{AB'} = \frac{1}{2^{1/2}} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned} \tag{10.8}$$

All the results of Chapter 1 can be immediately translated to curved space-time.

There is a one to two correspondence between tetrad transformations (10.3), (10.4) and spin transformations:

$$L^{\alpha}{}_{\beta}(x) \leftrightarrow \pm \lambda^{A}{}_{B}. \tag{10.7}$$

Under coordinate transformations, spinors remain invariant. We also have

$$\sigma_a{}^{A}{}_{C'}\sigma_b{}^{BC'} + \sigma_b{}^{A}{}_{C'}\sigma_a{}^{BC'} = g_{ab}\epsilon^{AB}, \tag{10.10}$$

$$g_{ab} \leftrightarrow \epsilon_{AB}\epsilon_{C'D'}, \tag{10.11}$$

$$T^{a}{}_{a} = T^{AB'}{}_{AB'}, \tag{10.12}$$

$$\eta_{abcd} = (-g)^{1/2}\epsilon_{abcd} \tag{10.13}$$

$$\leftrightarrow \eta_{AE'BF'CG'DH'} = i(\epsilon_{AC}\epsilon_{BD}\epsilon_{E'H'}\epsilon_{F'G'} - \epsilon_{AD}\epsilon_{BC}\epsilon_{E'G'}\epsilon_{F'H'}).$$

Exercise. Show that a null vector k^a ($k^a k_a = 0$) pointing into the future has the spin representation

$$k^a \leftrightarrow k_{AB'} = k_A \overline{k}_{B'}, \tag{10.14}$$

where the 1-spinor k_A is determined to within a phase transformation $k_A \rightarrow k_A e^{i\theta}$ with real θ. Find the spin representation of a null vector pointing into the past.

Bivectors. A bivector is a skew symmetric second order tensor $F_{mn} = -F_{nm}$. Its spin representation $F_{AC'BD'}$ can be split into a part[30] $F_{A(C'|B|D')}$ symmetric in C' and D', and a part skew symmetric in C' and D' which must clearly be proportional to $\epsilon_{C'D'}$:

$$F_{AC'BD'} = F_{A(C'|B|D')} + \epsilon_{C'D'}\Phi_{AB}, \qquad \Phi_{AB} = \frac{1}{2} F_{AC'B}{}^{C'}. \tag{10.15}$$

The skew symmetry of F_{mn} implies the symmetry of Φ_{AB} and the skew symmetry in A and B of $F_{A(C'|B|D')}$, which therefore must be proportional to ϵ_{AB}. Thus

$$F_{mn} \leftrightarrow F_{AC'BD'} = \epsilon_{AB}\psi_{C'D'} + \epsilon_{C'D'}\Phi_{AB},$$

$$\Phi_{AB} = \Phi_{BA} = \frac{1}{2} F_{AC'B}{}^{C'}, \qquad \psi_{C'D'} = \psi_{D'C'} = \frac{1}{2} F_{AC'}{}^{A}{}_{D'}. \tag{10.16}$$

[30] Suffixes between vertical lines such as B in $F_{A(C'|B|D')}$ are excluded from the symmetrization. The same applies to the case of skew symmetrization.

If F_{mn} is real, then it is easy to see (exercise) that $\psi_{C'D'} = \bar{\Phi}_{C'D'}$. Therefore the spin representation of a real bivector is

$$F_{mn} \leftrightarrow F_{AC'BD'} = \epsilon_{AB}\bar{\Phi}_{C'D'} + \epsilon_{C'D'}\Phi_{AB}, \qquad \Phi_{AB} = \Phi_{BA}. \tag{10.17}$$

Thus the real bivector F_{mn} and the symmetric spinor Φ_{AB} determine one another.

By Equation (10.13), the dual bivector $F^*{}_{mn} = \frac{1}{2}\eta_{mnrs}F^{rs}$ has the spin representation

$$F^*{}_{mn} \leftrightarrow F^*{}_{AC'BD'} = i(\epsilon_{AB}\bar{\Phi}_{C'D'} - \epsilon_{C'D'}\Phi_{AB}). \tag{10.18}$$

The complex bivector

$$F^+{}_{mn} = F_{mn} + iF^*{}_{mn} \tag{10.19}$$

is called self-dual because $(F^+{}_{mn})^* = -iF^+{}_{mn}$. It has the simple spin representation

$$F^+{}_{mn} \leftrightarrow F^+{}_{AC'BD'} = 2\epsilon_{C'D'}\Phi_{AB}. \tag{10.20}$$

Exercise. Show that the mapping

$$\Phi_{AB} \longrightarrow \Phi_{AB}e^{-i\theta} \tag{10.21}$$

corresponds to

$$F_{mn} \longrightarrow F_{mn}\cos\theta + F^*{}_{mn}\sin\theta. \tag{10.22}$$

This is called a *duality rotation.*

The expression $\Phi_{AB}\xi^A\xi^B$ is $(\xi^2)^2$ times a quadratic polynomial in the ratio ξ^1/ξ^2. It can therefore be factored:

$$\Phi_{AB}\xi^A\xi^B = (\xi^2)^2(k_1\xi^1/\xi^2 + k_2)(m_1\xi^1/\xi^2 + m_2) = k_A\xi^A m_B\xi^B, \tag{10.23}$$

where the spinors k_A and m_A are determined to within a constant of proportionality. Since Φ_{AB} is symmetric, it follows that

$$\Phi_{AB} = k_{(A}m_{B)}. \tag{10.24}$$

The null vectors $k^a \leftrightarrow k_A\bar{k}_{B'}$ and $m^a \leftrightarrow m_A\bar{m}_{B'}$ are called *principal null directions* of the bivector F_{mn}. If $k_A = \lambda m_A$, the principal null vector k^a is called *double*, F_{mn} is called *algebraically special* or *null*, and, by a suitable choice of proportionality factor, we can write $\Phi_{AB} = k_{(A}k_{B)}$. If k_A and m_A are not proportional, the principal null directions are called *simple* and F_{mn} is called *algebraically general.*

Exercise. Show that a principal null direction is characterized

by the equation

$$\Phi_{AB}k^A = \lambda k_B \Leftrightarrow k^A\Phi_{A[B}k_{E]} = 0, \tag{10.25}$$

that the null direction is simple if $\lambda \neq 0$, and double if $\lambda = 0$.

We shall now find tensor equivalents of the above criteria. We easily check that

$$F^+{}_{a[b}k_{c]}k^a \leftrightarrow 2k^A\Phi_{A[B}k_{E]}\bar{k}_{D'}\bar{k}_{F'}, \tag{10.26}$$

$$F^+{}_{ab}k^a \leftrightarrow k^A\Phi_{AB}\bar{k}_{C'}. \tag{10.27}$$

It follows that a simple principal null direction satisfies the equation

$$F^+{}_{a[b}k_{c]}k^a = 0, \tag{10.28}$$

but does not satisfy the stronger equation

$$F^+{}_{ab}k^a = 0. \tag{10.29}$$

A double principal null direction satisfies the stronger equation (10.29).

We summarize the algebraic classification of a real bivector in the following Table III:

TABLE III

Partition	*Name*	*Form of* Φ_{AB}	*Equation for* Φ_{AB}	*Equation for* $F^+{}_{mn}$
[11]	General	$\Phi_{AB} = k_{(A}m_{B)}$	$k^A\Phi_{AB} = \lambda k_B$	$F^+{}_{a[b}k_{c]}k^a = 0$
[2]	Null	$\Phi_{AB} = k_Ak_B$	$k^A\Phi_{AB} = 0$	$F^+{}_{ab}k^a = 0$

In this table, $\lambda \neq 0$, k_A and m_A are not proportional, and the partition gives the multiplicity of the principal null directions.

Exercises. 1. Show that a real bivector F_{ab} is null if and only if there exists a real vector k^a such that

$$F_{ab}k^a = 0, \qquad F^*{}_{ab}k^a = 0. \tag{10.30}$$

In particular, derive from these equations that k^a is null. Deduce that a discontinuity ΔF_{ab} of the electromagnetic field across a hypersurface must be null (cf. Chapter 7).

2. Show that F_{ab} is null if and only if

$$\Phi_{AB}\Phi^{AB} = 0 \Leftrightarrow F_{ab}F^{ab} = 0, \qquad F^*{}_{ab}F^{ab} = 0. \tag{10.31}$$

3. Let k^a be a constant null vector in Minkowski space. A plane electromagnetic wave travelling in the k^a direction has the form

$F_{mn} = k_{[m}v_{n]}$, where $v_n(u)$ is any vector function of the argument $u = k_a x^a$ subject to the restriction $k_a v^a = 0$, x^a being rectangular Cartesian coordinates. Check that F_{mn} satisfies Maxwell's equations, that it represents a plane wave, and that it is null.

Geometrical Interpretation of 1-*Spinors.* A 1-spinor k_A determines uniquely a real null bivector F_{mn} characterized by the symmetric spinor

$$\phi_{AB} = k_A k_B, \tag{10.32}$$

$$F_{mn} \leftrightarrow \epsilon_{AB}\bar{k}_{C'}\bar{k}_{D'} + \epsilon_{C'D'}k_A k_B. \tag{10.33}$$

Conversely F_{mn} or ϕ_{AB} determine k_A to within a sign.

Choose a spinor m_A which is linearly independent of k_A and normalize it so that

$$k_A m^A = 1 \Leftrightarrow \epsilon_{AB} = k_A m_B - k_B m_A. \tag{10.34}$$

The two spinors determine the real vectors

$$k^a \leftrightarrow k_A \bar{k}_{B'}, \tag{10.35}$$

$$v^a \leftrightarrow k_A \bar{m}_{B'} + m_A \bar{k}_{B'}. \tag{10.36}$$

It is easily checked that

$$k_a k^a = 0, \quad k_a v^a = 0, \quad v_a v^a = -2, \tag{10.37}$$

so that k^a is null and v^a is a space-like vector orthogonal to k^a.

If we substitute ϵ_{AB} from Equation (10.34) into Equation (10.33),

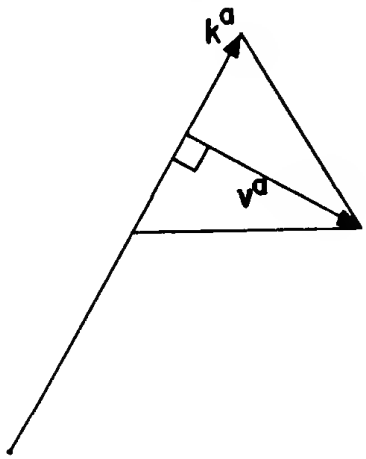


FIGURE 35

we find

$$F_{mn} = 2k_{[m}v_{n]}. \tag{10.38}$$

Thus a 1-spinor k_A can be represented geometrically by a *flag*, the flagpole being the null vector k^a and the position of the flag being determined by F_{mn} as shown in Figure 35. Conversely the flag represents either one of the two 1-spinors $\pm k_A$.

A continuous field of flags in a simply connected domain of space-time will, by continuity, determine the relative signs of the corresponding spinors. Thus a field of flags together with a single sign chosen at one point form the geometrical representation of a spinor field.

Exercises. 1. Show that m^A is determined by Equation (10.34) to within the change $m^A \to m^A + \lambda k^A$, and that v^a is determined to within the change $v^a \to v^a + (\lambda + \bar{\lambda})\, k^a$.

2. Show that under the phase change $k_A \to e^{i\theta}k_A$, we have $v^a \to v^a \cos 2\theta + w^a \sin 2\theta$, where $w^a \leftrightarrow i(k_A \bar{m}_{B'} - m_A \bar{k}_{B'})$ is a real vector orthogonal to k^a and v^a.

The Curvature Spinors. We shall now study the spin representation of the Riemann tensor

$$R_{abcd} \leftrightarrow R_{AE'BF'CG'DH'}. \tag{10.39}$$

The procedure is analogous to the bivector case but more complicated. We first split $R_{AE'BF'CG'DH'}$ into parts skew symmetric and symmetric in E' and F', using $R_{abcd} = -R_{bacd}$ and the reality of R_{abcd}. We obtain

$$R_{AE'BF'CG'DH'} = \epsilon_{E'F'}\alpha_{ABCG'DH'} + \epsilon_{AB}\bar{\alpha}_{E'F'G'CH'D}, \tag{10.40}$$

where $\alpha_{ABCG'DH'} = \alpha_{(AB)CG'DH'}$. Next we use $R_{abcd} = -R_{abdc}$, and obtain

$$\alpha_{ABCG'DH'} = \epsilon_{G'H'}\beta_{ABCD} + \epsilon_{CD}\phi_{ABG'H'}, \tag{10.41}$$

where $\beta_{ABCD} = \beta_{(AB)(CD)}$ and $\phi_{ABG'H'} = \phi_{(AB)(G'H')}$. The symmetry $R_{abcd} = R_{cdab}$ gives

$$\beta_{ABCD} = \beta_{(AB)(CD)} = \beta_{CDAB}, \tag{10.42}$$

$$\phi_{ABG'H'} = \phi_{(AB)(G'H')} = \bar{\phi}_{G'H'AB}. \tag{10.43}$$

The final symmetry $R_{abcd}\eta^{mbcd} = 0$ of the Riemann tensor gives, using Equation (10.13),

$$\epsilon_{AM}\bar{\beta}_{E'F'N'}{}^{F'} - \epsilon_{E'N'}\beta_{ABM}{}^{B} = 0, \tag{10.44}$$

from which we can derive

$$\beta_{ABCD} = \psi_{ABCD} + \frac{1}{6}(\epsilon_{AD}\epsilon_{BC} + \epsilon_{AC}\epsilon_{BD})\Lambda, \tag{10.45}$$

where

$$\psi_{ABCD} = \beta_{(ABCD)}, \qquad \Lambda = \bar{\Lambda} = \beta_{AB}{}^{AB}. \tag{10.46}$$

Combining these results we obtain the spin representation of the Riemann tensor:

$$\begin{aligned} R_{AE'BF'CG'DH'} &= \epsilon_{AB}\epsilon_{CD}\bar{\psi}_{E'F'G'H'} + \epsilon_{E'F'}\epsilon_{G'H'}\psi_{ABCD} \\ &\quad + \epsilon_{AB}\epsilon_{G'H'}\phi_{CDE'F'} + \epsilon_{E'F'}\epsilon_{CD}\phi_{ABG'H'} \\ &\quad + \frac{1}{3}(\epsilon_{AC}\epsilon_{BD}\epsilon_{E'G'}\epsilon_{F'H'} - \epsilon_{AD}\epsilon_{BC}\epsilon_{E'H'}\epsilon_{F'G'})\Lambda, \end{aligned} \tag{10.47}$$

$$\psi_{ABCD} = \psi_{(ABCD)}, \quad \phi_{ABG'H'} = \phi_{(AB)(G'H')} = \bar{\phi}_{G'H'AB}, \quad \bar{\Lambda} = \Lambda.$$

We shall now find the tensor equivalents of the curvature spinors ψ_{ABCD}, $\phi_{ABG'H'}$ and Λ. We contract to obtain the Ricci tensor

$$R_{bc} = R^{a}{}_{bca} \leftrightarrow R_{BF'CG'} = 2\phi_{BCF'G'} - \epsilon_{BC}\epsilon_{F'G'}\Lambda. \tag{10.48}$$

Contracting again, we find the curvature scalar

$$R = R^{b}{}_{b} = -4\Lambda. \tag{10.49}$$

Combining the last two equations, we see that $\phi_{CBG'F'}$ represents the trace-free Ricci tensor according to the correspondence

$$P_{bc} = R_{bc} - \frac{1}{4}g_{bc}R \leftrightarrow P_{BF'CG'} = 2\phi_{BCF'G'}. \tag{10.50}$$

Let us consider the Weyl or conformal tensor

$$C_{abcd} = R_{abcd} + g_{a[c}R_{d]b} + R_{a[c}g_{d]b} - \frac{1}{3}g_{a[c}g_{d]b}R. \tag{10.51}$$

This tensor has all the algebraic symmetries of the curvature tensor and moreover satisfies $C^{a}{}_{bca} = 0$ (exercise). It follows that C_{abcd} can be represented as in Equation (10.47), and the trace-

free condition implies that only the first two terms on the right-hand side survive:

$$(10.52)\quad C_{abcd} \leftrightarrow C_{AE'BF'CG'DH'} = \epsilon_{AB}\epsilon_{CD}\bar{\psi}_{E'F'G'H'} + \epsilon_{E'F'}\epsilon_{G'H'}\psi_{ABCD}.$$

Because R_{abcd} and C_{abcd} differ by terms that only involve the Ricci tensor, it is clear that ψ_{ABCD} is the same in Equations (10.47) and (10.52). Thus the three curvature spinors represent respectively the conformal tensor, the trace-free Ricci tensor and the curvature invariant.

We define the left and right duals of the conformal tensor by

$$(10.53)\qquad {}^*C_{abcd} = \frac{1}{2}\eta_{rsab}C^{rs}{}_{cd}, \quad C^*{}_{abcd} = \frac{1}{2}\eta_{rscd}C_{ab}{}^{rs}.$$

Applying the result of Equation (10.18) to either the first or the second pair of suffixes of C_{abcd}, we immediately find that the two duals are equal and have the spin representation

$$(10.54)\quad {}^*C_{abcd} = C^*{}_{abcd} \leftrightarrow i(\epsilon_{AB}\epsilon_{CD}\bar{\psi}_{E'F'G'H'} - \epsilon_{E'F'}\epsilon_{G'H}\psi_{ABCD}).$$

We form the complex conformal tensor

$$(10.55)\qquad C^+{}_{abcd} = C_{abcd} + iC^*{}_{abcd} \leftrightarrow 2\epsilon_{E'F'}\epsilon_{G'H'}\psi_{ABCD}.$$

The Petrov-Pirani Classification.[31] This algebraic classification of the conformal tensor is completely analogous to the classification of bivectors. When the gravitational vacuum equations $R_{mn} = 0$ are satisfied, $C_{abcd} = R_{abcd}$. Therefore the classification applies automatically to the vacuum curvature tensor.

By the same argument which led to Equation (10.24), we can reduce the symmetric spinor ψ_{ABCD} to the form

$$(10.56)\qquad \psi_{ABCD} = k_{(A}m_Br_Cs_{D)},$$

where the 1-spinors are determined to within constants of proportionality. The four null directions $k^a \leftrightarrow k_A\bar{k}_{B'}$, etc., are called *principal null directions* or *Debever-Penrose directions*. If all these directions are distinct, we say that C_{abcd} is *algebraically general*, with the *partition* [1111]. When some of the directions coincide, we say that C_{abcd} is *algebraically special*. We can then have the

[31] A. Z. Petrov, Sci. Not. Kazan State University **114** (1954), 55; F. A. E. Pirani, Phys. Rev. **105** (1957), 1089.

TABLE IV

(1)	(2) Partition	(3) Name	(4) Symbol	(5) Form of ψ_{ABCD}	(6) Equation for ψ_{ABCD}	(7) Equation for $C^{+}{}_{abcd}$
Algebraically general	[1111]	Petrov type I nondegenerate	I	$k_{(A}m_Br_Cs_{D)}$	$\psi_{ABCD}k^Bk^Ck^D = \lambda k_A$	$k_{[r}C^{+}{}_{a]bc[d}k_{s]}k^bk^c = 0$
Algebraically special	[211]	Petrov type II nondegenerate	II	$k_{(A}k_Br_Cs_{D)}$	$\psi_{ABCD}k^Ck^D = \lambda k_Ak_B$	$C^{+}{}_{abc[d}k_{s]}k^bk^c = 0$
	[22]	Petrov type I degenerate	D	$k_{(A}k_Br_Cr_{D)}$		
	[31]	Petrov type III	III	$k_{(A}k_Bk_Cs_{D)}$	$\psi_{ABCD}k^D = \lambda k_Ak_Bk_C$	$C^{+}{}_{abc[d}k_{s]}k^c = 0$
	[4]	Petrov type II degenerate or null	N	$k_Ak_Bk_Ck_D$	$\psi_{ABCD} = k_Ak_Bk_Ck_D$	$C^{+}{}_{abcd}k^c = 0$
$C_{abcd} = 0$		Conformally flat	O	0	—	$C^{+}{}_{abcd} = 0$

different possible partitions shown in column (2) of Table IV, e.g., [211] indicates that two principal null directions coincide and that the other two are distinct.

Column (5) of the table gives expressions for ψ_{ABCD} in the different cases. Here it is assumed that 1-spinors with different kernel letters are not proportional.

In Column (6) it is assumed that $\lambda \neq 0$. The successive equations then state that k^A corresponds to a simple, double, triple or quadruple principal null direction (exercise).

Column (7) gives the corresponding tensor equations: k^a is a simple principal null direction if it satisfies the first equation but not the second; it is a double principal null direction if it satisfies the second equation but not the third; etc. This is proved by equations analogous to (10.26) and (10.27). For example, with $k_s \leftrightarrow k_S \bar{k}_{V'}$, we obtain from Equation (10.55)

$$C^+{}_{abc[d}k_{s]}k^c \leftrightarrow 2\epsilon_{E'F'}\psi_{ABC[D}k_{S]}k^C\bar{k}_{H'}\bar{k}_{V'}, \tag{10.57}$$

$$C^+{}_{abcd}k^c \leftrightarrow 2\epsilon_{E'F'}\psi_{ABCD}k^C\bar{k}_{H'}. \tag{10.58}$$

We therefore have the following implications

$$\begin{aligned}\psi_{ABCD} \propto k_{(A}k_Bk_Cs_{D)} &\Leftrightarrow \psi_{ABCD}k^Ck^D = 0 \Leftrightarrow \psi_{ABC[D}k_{S]}k^C = 0 \\ &\Leftrightarrow C^+{}_{abc[d}k_{s]}k^c = 0,\end{aligned} \tag{10.59}$$

$$\psi_{ABCD} \propto k_Ak_Bk_Ck_D \Leftrightarrow \psi_{ABCD}k^C = 0 \Leftrightarrow C^+{}_{abcd}k^c = 0. \tag{10.60}$$

If we wish Equations (10.59) to hold with k_A and s_A not proportional, then the necessary and sufficient condition is $C^+{}_{abc[d}k_{s]}k^c = 0$, $C^+{}_{abcd}k^c \neq 0$.

Exercises. 1. Show that each of the equations in Column (7) of Table IV implies $k_ak^a = 0$, so that this need not be assumed separately.

2. Given that $k_ak^a = 0$, show that each equation in Column (7) of Table IV is equivalent to the same equation with $C^+{}_{abcd}$ replaced by C_{abcd}. Deduce that a discontinuity ΔR_{abcd} of the vacuum curvature tensor across a hypersurface is of Petrov type N.

THE UNIVERSITY OF TEXAS
AUSTIN, TEXAS

L. I. Schiff

Comparison of Theory and Observation in General Relativity*

These lectures have a twofold purpose: (A) to stress the care that must be taken in comparing predictions of general relativity theory with observations; and (B) to describe a quantitative method for determining the extent to which one particular but important aspect of the theory is being tested.

Point (A) might seem at first to be a trivial one, since a comparison between theory and experiment in any area of physics is essentially a comparison between two sets of numbers that represent the same physical quantity, one set calculated from the theory and the other from the observations. However it is not always easy, when dealing with general relativity, to know what is measurable and what is an artifact of the choice of coordinate system, which in this theory is completely arbitrary. For example, one of the early studies of the gyroscope argued that the magnitude of its angular momentum would vary as it moves through a gravitational field, whereas in actuality only the direction of this vector changes. From the point of view of an experimenter planning to observe a

* Work supported in part by the U.S. Air Force through Air Force Office of Scientific Research Contract AF 49(638)-1389.

gyroscope, it makes a great deal of difference which effect is expected to occur, since a small precession of the spin axis is far more difficult to measure than a correspondingly small change in rotation rate, and requires a quite different technique.

A convenient heuristic device is available to the physicist who is planning such a comparison between theory and observation. This consists in exploiting the coordinate invariance of the physically observable quantities by calculating them in different coordinate systems and (when appropriate) with different supplementary conditions. The quantities calculated in this way are of course all obtainable from each other by suitable transformations. However the process of reconciling them without using these transformations forces the investigator to discard coordinate-dependent artifacts, and retain only physically meaningful relations between measurable quantities. This procedure will be illustrated later in these lectures.

The test method referred to in point (B) goes back at least to 1922. At that time, Eddington [1] remarked that different terms in the expression for the metric tensor are responsible for different tests of the theory, provided that the connection between the metric and the observable quantities remains the same as in general relativity. This corresponds to modifying Einstein's theory in a special but nevertheless instructive and interesting way. This program has been extended by Robertson [2], and the present discussion carries it further.

We write the metric in the most general static spherically symmetric form, as representative of the gravitational field about an assumed stationary, nonrotating, spherically symmetric sun or earth. In units such that the Newtonian gravitational constant and the speed of light are equal to unity, this is [2]

$$ds^2 = [1 - (2\alpha m/r) + (2\beta m^2/r^2) + \cdots]dt^2 - [1 + (2\gamma m/r) + \cdots](dx^2 + dy^2 + dz^2),$$

where $r^2 = x^2 + y^2 + z^2$, m is the mass of the central attracting body, and $\alpha, \beta, \gamma, \cdots$ are numerical coefficients that are all equal to unity in Einstein's theory. We also assume that the equations of motion of matter and electromagnetic radiation are given in terms of this metric tensor by Einstein's equations. In this way we can find the dependence of the calculated quantities on $\alpha, \beta, \gamma, \cdots$, which in turn measure the time or space distortions that are linear or

nonlinear in m and therefore in the gravitational field strength. While this procedure is to some extent arbitrary, it is well defined and provides a common basis for comparison of various possible experimental tests of general relativity theory.

Einstein's three tests. The three tests originally proposed by Einstein do not present any difficulty in connection with point (A). The gravitational red shift is observed by comparison of two infinitesimal time intervals, frequencies or spectral wavelengths at the same space-time point, and there can be no ambiguity as to what is measured or what is to be calculated. The gravitational deflection of starlight by the sun and the advance of the perihelion of the orbit of an inner planet are both observed through measurement of angles. Since the gravitational field of the sun is believed to be spherically symmetric to very high accuracy, and the "fixed stars" provide a convenient background against which to measure angles, such observations are again unambiguous although they do of course require extraordinary care [3].

With regard to point (B), the dependence of the theoretical predictions for these three effects on the numerical values of α, β, and γ has been known for a long time [1] and has recently been summarized by Robertson [2]. We note first that with the conventional definition of the gravitational constant, we must choose $\alpha = +1$ in order for planetary orbits to agree with Newtonian theory in lowest order. Then since the first-order gravitational red shift depends only on the α term in the metric, no new information concerning the metric is provided by such a measurement: This conclusion is equivalent to the statement that the first-order red shift follows correctly from any theory of gravitation that is consistent with the equivalence principle [4].

With this choice of α, the deflection of starlight turns out to be proportional to $1 + \gamma$, and the advance of the perihelion of a planetary orbit is proportional to $2(1 + \gamma) - \beta$. Astronomical observations show that $1 + \gamma$ is equal to 2 with an uncertainty of roughly 20%, and that $2(1 + \gamma) - \beta$ is equal to 3 with an uncertainty of about 1%. Thus while we know that $\beta = 2\gamma - 1$ with high precision, γ is poorly determined.

Two of the most characteristic and interesting aspects of Einstein's theory are its nonlinearity, and its prediction that the gravitational field produced by matter in motion is different from that produced

by the same matter at rest. We return to the second aspect in the next section, and point out here that the β term in the metric is the simplest manifestation of the nonlinearity of the theory. It is important, therefore, to determine β as accurately as possible. The preceding paragraph shows that at the present time this requires more precise knowledge of γ, which intrinsically is less interesting than β.

Gyroscope precession. Three other possible measurements will be discussed in this and the following sections: the precession of the spin axis of a spherical gyroscope that moves through a gravitational field [5], the second-order gravitational red shift, and the transit time of radar signals reflected from an inner planet [6].

The precession of the spin axis of a gyroscope can be calculated by making use of Fermi-Walker transport [7] of a coordinate system. However such a calculation rests on the geometrical assumption that the spin angular momentum has a fixed magnitude and orientation in this coordinate system. It is preferable to perform the calculation dynamically rather than geometrically; the equations of motion were first obtained in this way by Papapetrou [8], who used the method of Fock [9].

These equations describe the changes in the angular momentum tensor S^{ij}, where the ordinary angular momentum vector is given by

$$\mathbf{S} = (S_x, S_y, S_z) = (S^{23}, S^{31}, S^{12}).$$

The other three components S^{0i} of this antisymmetrical tensor (x^0 is the time coordinate) must be disposed of by a supplementary condition. Corinaldesi and Papapetrou [10] assumed that these components are zero in the rest frame of the central attracting body, and solved the equations of motion with the standard form of the Schwarzschild metric [11]. Their result, to first order in m, is

$$d\mathbf{S}/dt = (m/r^3)\,[2\mathbf{S}(\mathbf{r}\cdot\mathbf{v}) + 2\mathbf{v}(\mathbf{r}\cdot\mathbf{S}) - \mathbf{r}(\mathbf{v}\cdot\mathbf{S}) - (3\mathbf{r}/r^2)(\mathbf{r}\cdot\mathbf{v})(\mathbf{r}\cdot\mathbf{S})\,],$$

where $\mathbf{v} = d\mathbf{r}/dt$. It is independent of whether or not a nongravitational force $\mathbf{F}$ is applied to the center of mass of the gyroscope. On this basis they argued that since $d\mathbf{S}/dt$ has a component parallel to $\mathbf{S}$, its magnitude will change, and hence the spin rotation frequency might be expected to change.

Their argument can be tested by making use of point (A) [5]. The calculation is next performed with the isotropic form of the Schwarzschild metric [11] and the same supplementary condition to obtain

$$d\mathbf{S}/dt = (m/r^3)\,[3\mathbf{S}(\mathbf{r}\cdot\mathbf{v}) + \mathbf{v}(\mathbf{r}\cdot\mathbf{S}) - 2\mathbf{r}(\mathbf{v}\cdot\mathbf{S})].$$

As a further application of point (A), we then use a different supplementary condition, introduced by Pirani [12]:

$$S^{ij}u_j = 0, \qquad u^i = dx^i/ds.$$

This leads to

$$d\mathbf{S}/dt = (3m/r^3)\,[\mathbf{v}(\mathbf{r}\cdot\mathbf{S}) - (\mathbf{r}/r^2)(\mathbf{r}\cdot\mathbf{v})(\mathbf{r}\cdot\mathbf{S})] + \mathbf{S}(\mathbf{v}\cdot\mathbf{f}) - \mathbf{f}(\mathbf{v}\cdot\mathbf{S})$$

with the standard form of the metric, and to

$$d\mathbf{S}/dt = (m/r^3)\,[\mathbf{S}(\mathbf{r}\cdot\mathbf{v}) + 2\mathbf{v}(\mathbf{r}\cdot\mathbf{S}) - \mathbf{r}(\mathbf{v}\cdot\mathbf{S})] + \mathbf{S}(\mathbf{v}\cdot\mathbf{f}) - \mathbf{f}(\mathbf{v}\cdot\mathbf{S})$$

with the isotropic form, where $\mathbf{f}$ is equal to $\mathbf{F}$ divided by the rest mass of the gyroscope.

These four expressions for $d\mathbf{S}/dt$ may be reconciled by recognizing that measurements on the gyroscope are in practice made by a co-moving observer, who then transmits his results to the outside world. If for example the gyroscope is in an earth satellite, it may be observed by a human being who travels with it, or by a device which telemeters information to the ground. The transformation from $\mathbf{S}$ to $\mathbf{S}_0$, which is the angular momentum vector measured in the rest frame of the gyroscope, consists of two parts: a co-ordinate transformation involving changes of order m/r, and a Lorentz transformation involving changes of order v^2. These transformations depend on the metric and supplementary condition that are employed. In addition, the point at which $\mathbf{F}$ is applied depends on the choice of supplementary condition, so that allowance must be made for a possible spurious torque. When these changes are taken into account, all four expressions become the same to first order [5]:

$$d\mathbf{S}_0/dt = \Omega \times \mathbf{S}_0, \qquad \Omega = (1/2)(\mathbf{f}\times\mathbf{v}) + (3m/2r^3)(\mathbf{r}\times\mathbf{v}).$$

Thus the magnitude of $\mathbf{S}_0$, and hence the spin rotation rate, does not change. The first term in Ω is the Thomas precession [13], which

is a special relativistic rather than a general relativistic effect since it does not involve m. The second term is a true gravitational effect, and is often called the geodetic precession; the contribution of this term when integrated over an orbital period was first discovered by deSitter [14].

Now that the four expressions for $d\mathbf{S}/dt$ have been reconciled, it is a relatively easy matter to calculate the effect of replacing the Schwarzschild metric by the more general form introduced in connection with point (B). The Thomas precession is, of course, not changed, since it is obtained even when m is zero. However the factor 3 in the expression for the geodetic precession now becomes $1 + 2\gamma$ [2], [15], so that this experiment can provide an independent measure of the value of γ.

The gyroscope experiment can also make possible a measurement of the difference between the gravitational field of the rotating and nonrotating earth. When earth rotation is included, the expression for Ω contains the additional term

$$(I/r^3)\,[(3\mathbf{r}/r^2)(\boldsymbol{\omega}\cdot\mathbf{r}) - \boldsymbol{\omega}],$$

where I is the moment of inertia and $\boldsymbol{\omega}$ the angular velocity vector of the earth [5]. This term is typically about two percent of the geodetic precession for a satellite-borne gyroscope, and its measurement would constitute the only feasible verification thus far proposed of the mass-current effect predicted by general relativity theory.

Second-order red shift. No one has ever seriously proposed that the gravitational red shift could be measured to second order in m, and this is not the intention here. Nevertheless, the analysis of such a possible measurement provides an interesting and instructive example of the application of points (A) and (B). We imagine that a standard clock (or nucleus) at $r_1 = R$ transmits a periodic signal upward along a radial line to $r_2 = R + h$, and ask for the ratio of the period as measured at r_2 to the period of an identical clock located at r_2. This ratio is $[g_{00}(r_2)/g_{00}(r_1)]^{1/2}$, which in the standard form of the Schwarzschild metric is equal to

$$\left[\frac{1-(2m/r_2)}{1-(2m/r_1)}\right]^{1/2} \cong 1 + \frac{mh}{R^2}\left[1 - \frac{h}{R} + \left(\frac{h}{R}\right)^2 + \frac{2m}{R}\right].$$

The terms kept in the expansion are consistent with a terrestrial

experiment for which $h = 60$ meters, in which case $m/R \sim 10^{-9}$ and $h/R \sim 10^{-5}$.

This result suggests that a measurement of the second-order red shift would verify the nonlinearity of Einstein's equations, since it contains a term proportional to m^2. However consideration of point (A) leads us to do the calculation also with the isotropic metric, which gives

$$\frac{[1-(m/2r_2)][1+(m/2r_1)]}{[1+(m/2r_2)][1-(m/2r_1)]} \cong 1+\frac{mh}{R^2}\left[1-\frac{h}{R}+\left(\frac{h}{R}\right)^2\right].$$

In order to reconcile these two results, we must consider the way in which the quantities that enter into the first-order expression mh/R^2 might actually be measured. A natural way in which to measure h is to use standard measuring rods; alternatively, the transit time of a light signal from r_1 that is reflected back at r_2 might be compared with a standard clock. It is found in both cases that the proper height difference l is given to sufficient accuracy in terms of the coordinate distance h by $l \cong h[1+(m/R)]$, when either metric is used. Thus the discrepancy is not resolved by the substitution of l for h.

Next we consider the quantity m/R^2, which to first order is the acceleration of gravity g at the point $r_1 = R$. Now neither m nor R is subject to direct measurement; this is unimportant for the correction terms obtained above, for which the usual Newtonian values can be used. However values that are themselves correct through first order must be inserted into the coefficient mh/R^2. We therefore consider how the combination m/R^2 or g might be measured directly. Two general methods suggest themselves: a static weighing experiment and a measurement of acceleration in free fall. An object may be weighed by balancing the gravitational pull of the earth against, say, a spring, and then calibrating the spring by using Newton's second law of motion to convert its force into an acceleration. Alternatively, the gravitational pull of the earth may be calibrated against the centrifugal force arising from the rotation of the earth by measurement of the angle between the true and apparent verticals [16]. The free fall measurement might be performed by dropping an object from r_2 and measuring the time interval until a light signal emitted on impact at r_1 returns to r_2; alternatively, the object could be projected upward from

r_1 in such a way that it just reaches r_2, and the time interval measured until return to r_1.

The free fall measurement seems to be the most straightforward, both theoretically and experimentally. Integration of the equation of motion leads to the following expression for m/R^2 in terms of the one-way coordinate time interval t,

$$m/R^2 = (2h/t^2)\,[1 + (4h/3R) + (31h^2/90R^2) + (2m/R)]$$

when the standard form of the metric is used, and to

$$m/R^2 = (2h/t^2)\,[1 + (4h/3R) + (31h^2/90R^2) + (4m/R)]$$

when the isotropic form is used. Thus substitution into the earlier expressions for the clock period ratio leads to

$$1 + (2h^2/t^2)\,[1 + (h/3R) + (h^2/90R^2) + (4m/R)]$$

for both forms of the metric. It is still necessary to replace h by the proper height difference l, and t by the proper time interval τ associated with free fall as measured, say at r_1. The pertinent relations are, to first order:

$$l \cong h\,[1 + (m/R)], \qquad \tau \cong t\,[1 - (m/R)],$$

for both metrics. Thus we finally obtain for the period ratio

$$1 + (2l^2/\tau^2)\,[1 + (l/3R) + (l^2/90R^2)].$$

This result is not only independent of the coordinate-system, as it must be, but contains no term nonlinear in m.

Now that the considerations of point (A) have suggested how the clock period ratio can be expressed in terms of measurable quantities, the entire calculation can be done over again with the more general form of metric introduced in connection with point (B). The result is that the gravitational red shift, through terms of second order in m, is independent of both β and γ! This unexpected conclusion makes the experiment uninteresting as well as infeasible.

Planetary radar reflection. Shapiro [6] has recently proposed that the transit time of radar signals reflected from an inner planet be measured in order to provide another test of general relativity theory. In Figure 1, the sun is at the origin, the earth has instantane-

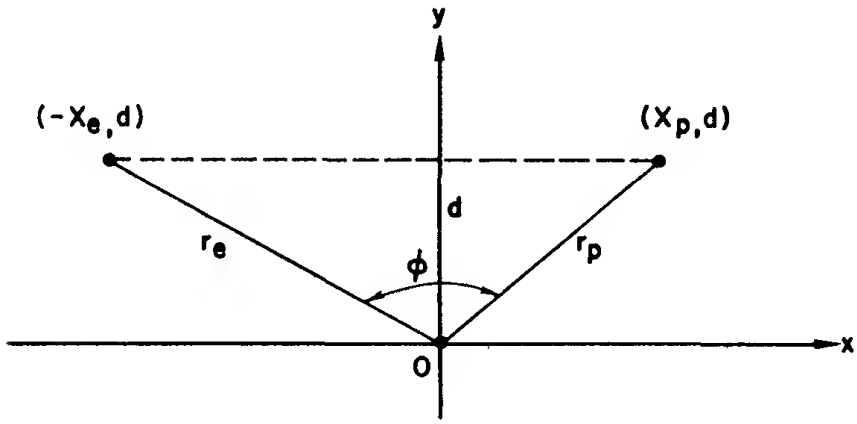


FIGURE 1

ous rectangular coordinates $(-x_e, d)$, and the planet (Venus or Mercury) has coordinates (x_p, d). A radar signal is sent from the earth to the planet and back, and the round trip transit time is measured. To zero order in the mass m of the sun, this time interval is $2(x_e + x_p)$, and an additional contribution of first order in m is expected on the basis of general relativity theory. We follow Shapiro in neglecting the motions of the earth and the planet during an observation, and also in neglecting the departure of the radar path from the straight line $y = d$. Path curvature is easily seen to produce a contribution of order m^2, which is negligibly small.

It is convenient to work entirely in terms of a universal coordinate time, which is applicable to all parts of the system since the metric is static. The final results must of course be converted to proper time as measured on the earth, but this common transformation will not affect the relationships between measured quantities that are derived below.

The coordinate transit time t may be computed by integration of the equations of motion of the electromagnetic signal, or much more easily by integrating the velocity equation

$$dx/dt = (-g_{00}/g_{11})^{1/2},$$

which is obtained from the metric by putting $ds = dy = dz = 0$. The result for the standard form of the Schwarzschild metric is

$$t_S = 2(x_e + x_p) + 4m \ln[(r_p + x_p)/(r_e - x_e)] - 2m[(x_e/r_e) + (x_p/r_p)],$$

which is equivalent to that obtained in [6]. Consideration of point (A) now leads us to repeat the calculation with the isotropic form of the metric; this gives

$$t_I = 2(x_e + x_p) + 4m \ln [(r_p + x_p)/(r_e - x_e)].$$

In order to reconcile these two expressions, we must consider the way in which the zero order distance $x_e + x_p$ might actually be measured. It is clearly impossible to use standard measuring rods, and light signals are inappropriate in the present situation. A plausible way in which to proceed is first to express $x_e + x_p$ in terms of r_e, r_p, and ϕ (see Figure 1); then since r_e and r_p are not directly measurable, they may be expressed in terms of the orbital periods, eccentricities, and elapsed times since perihelion. The angle ϕ can also be expressed in terms of these quantities, together with the elapsed time since conjunction. In this way, $x_e + x_p$ is ultimately expressible in terms of measured times; as remarked above, it is convenient to choose these to be universal coordinate times, but they are readily convertible to earth proper times.

It is sufficient for the present discussion to approximate the orbits as circles. Solution of the equations of motion then shows that the standard metric leads in first order to the same relation between r_p and the orbital period T_p as is obtained in Newtonian theory (Kepler's third law):

$$r_p = (mT_p^2/4\pi^2)^{1/3}.$$

The corresponding relation obtained with the isotropic metric is

$$r_p = (mT_p^2/4\pi^2)^{1/3} - m.$$

Now the distance $x_e + x_p$ is given in terms of r_e, r_p, and ϕ by the law of cosines, which is valid here since r is defined to be $(x^2 + y^2 + z^2)^{1/2}$:

$$x_e + x_p = (r_e^2 + r_p^2 - 2r_e r_p \cos\phi)^{1/2}.$$

The transition from the isotropic to the standard metric is accomplished by replacing r_e by $r_e - m$ and r_p by $r_p - m$; this has the effect of changing $x_e + x_p$ into

$$x_e + x_p - m[(x_e/r_e) + (x_p/r_p)].$$

Substitution into the above expression for t_I then shows that it becomes the same as t_S.

We see then that t_S is the coordinate transit time expected from the Einstein theory, provided that the zero-order term is computed from ϕ and the orbital periods in accordance with Newtonian theory.

Turning now to point (B), it can be shown that the generalized metric leads to the following expression for t:

$$2(x_e + x_p) + 2(1+\gamma)m \ln[(r_p + x_p)/(r_e - x_e)] - \frac{2(\gamma + 2\beta)}{3} m[(x_e/r_e) + (x_p/r_p)].$$

Thus t depends not only on γ, as expected in an electromagnetic propagation experiment, but also on β. It is true that the β dependence enters indirectly, through the calibration of the zero-order distance $x_e + x_p$ in terms of the orbital periods. Nevertheless, this result endows the proposed radar reflection experiment with the added significance that it and the planetary perihelion advance are the only observations that are sensitive to a nonlinear term in Einstein's theory.

It is a pleasure to thank Mr. Dennis K. Ross for several informative discussions of the material presented in the last two sections. A fuller account of the planetary radar reflection analysis will be published elsewhere in collaboration with Mr. Ross.

Note added in proof. Since these lectures were prepared, the paper referred to in the last paragraph has appeared [17], and Shapiro has published a critique of this paper [18] which suggests that the presence of the β-dependent term in the last equation above may be masked by uncertainties in the effective planetary radii. It seems to us that the possibility of such masking adds to the importance of experiments with instrumented space probes that have been proposed recently [19]. These experiments would make use of solar or planetary orbiters that carry radar transponders and present targets that are much better defined than the planets themselves.

The attention of the interested reader is called to two recent reviews of the experimental aspects of general relativity [20], [21].

References

1. A. S. Eddington, *The mathematical theory of relativity,* Cambridge Univ. Press, New York, 1957, p. 105.

2. H. P. Robertson, *Space age astronomy*, A. J. Deutsch and W. B. Klemperer, eds., Academic Press, New York, 1962, p. 228.

3. G. M. Clemence, Rev. Modern Phys. **19** (1947), 361.

4. A. Einstein, Ann. Physik **35** (1911), 898.

5. L. I. Schiff, Proc. Nat. Acad. Sci. U.S.A. **46** (1960), 871.

6. I. I. Shapiro, Phys. Rev. Lett. **13** (1964), 789; Lincoln Laboratory Report No. 368, December, 1964 (unpublished).

7. J. L. Synge, *Relativity: the general theory*, North-Holland, Amsterdam, 1960, p. 13.

8. A. Papapetrou, Proc. Roy. Soc. Ser. A **209** (1951), 248.

9. V. A. Fock, J. Physics U.S.S.R. **1** (1939), 81.

10. E. Corinaldesi and A. Papapetrou, Proc. Roy. Soc. Ser. A **209** (1951), 259.

11. R. Adler, M. Bazin and M. Schiffer, *Introduction to general relativity*, McGraw-Hill, New York, 1965, pp. 173, 176.

12. F. A. E. Pirani, Acta Phys. Polon. **15** (1956), 389.

13. L. H. Thomas, Philos. Mag. (7) **3** (1927), 1.

14. W. deSitter, Monthly Notices Roy. Astronom. Soc. **77** (1916), 155, 481.

15. L. I. Schiff, *Proceedings on theory of gravitation*, L. Infeld, ed., Gauthier-Villars, Paris; PWN-Polish Scientific Publishers, Warsaw, 1964, p. 71.

16. This is similar to the concept behind the experiment of R. V. Eötvös, D. Pekár, and E. Fekete, Ann. Physik. **68** (1922), 11.

17. D. K. Ross and L. I. Schiff, Phys. Rev. **141** (1966), 1215.

18. I. I. Shapiro, Phys. Rev. **145** (1966), 1005.

19. V. R. Eshleman, private communication.

20. M.-A. Tonnelat, *Les vérifications expérimentales de la relativité générale*, Masson et Cie, Paris, 1964.

21. V. L. Ginzburg, Astronautica Acta, (to appear).

INSTITUTE OF THEORETICAL PHYSICS, STANFORD UNIVERSITY
STANFORD, CALIFORNIA

Freeman J. Dyson

Experimental Tests of General Relativity. Radar

I. **Introduction.** There is not really enough material for an hour's lecture on the Radar Test of general relativity (which is described by I. I. Shapiro, Phys. Rev. Lett. **13** (1964), 789-791). So I will talk about radar observations of the planets in general, with the test of general relativity as one particular application.

One reason for being interested in radar astronomy is the very rapid rate of growth of electrical engineering as compared with optical engineering. To see how radar performance has improved over the years, let us take as a figure of merit the ratio

$$A = \frac{\text{range}^4}{\text{detectable cross section}},$$

since the least cross section visible at range R varies with R^4 for any given radar system.

In 1940 at the beginning of World War 2 we could barely detect an airplane with $10\,\text{m}^2$ cross section at a distance of $100\,\text{km} = 10^5\text{m}$. Hence

$$(1940) \qquad A = 10^{19}\text{m}^2.$$

In 1960 the Lincoln Laboratory people detected Venus with a

cross section of about 10^8km^2 at a distance of 10^8km. Thus

(1960) $$A = 10^{30} \text{m}^2.$$

This year the observations at Arecibo and at Goldstone have detected not merely the planet Venus as a whole but continents (or mountain chains) upon its surface. This means a cross section of 10^6km^2 rather than 10^8km^2, so we have

(1965) $$A = 10^{32} \text{m}^2.$$

We see that the performance of radar has improved by one power of 10 every 2 years, quite consistently from 1940 to 1965. The improvement is composed about equally of four ingredients,

1. Higher-gain antennas,
2. Higher-power transmitters,
3. Lower-noise receivers,
4. More sophisticated data-processing.

All four items are capable of further improvement, and so we may expect the factor 10 per 2 years improvement to continue for some time.

Contrast this with optical astronomy. Herschel in 1765 could see things down to about magnitude 12, and now we are at about magnitude 22, an improvement by 4 powers of 10 in 200 years, or one power of 10 in 50 years. Image intensifiers may improve the rate of progress in optical engineering, but clearly we cannot expect anything as spectacular as the radar people have achieved.

II. **Current work in planetary radar.** There are at present three main groups observing planets with radar.

1. Arecibo Ionospheric Observatory. 430 Mc which means 75 cm. Pettengill in charge.

2. Goldstone (JPL). 2400 Mc or 12 cm. Goldstein, Carpenter and other people.

3. Lincoln Laboratory did the original observation of Venus at 50 Mc. Now has built the Haystack radar, 8000 Mc or 4 cm. This will be ready for doing planetary work in about a year. I. I. Shapiro in charge.

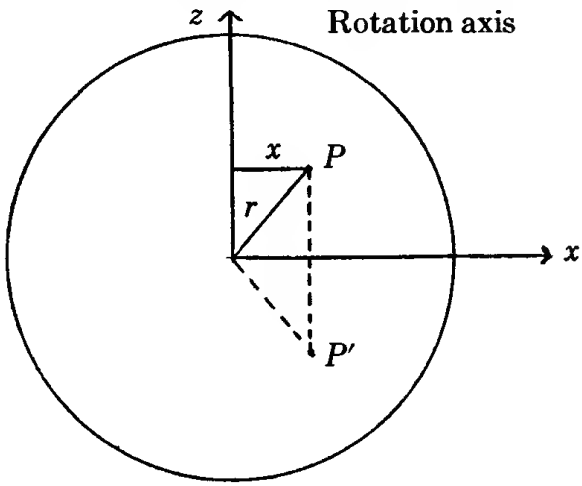
Goldstone has at present an 84 foot dish; a 210 foot dish, also at 2400 Mc, is almost finished.

There are two things of extreme importance which are done with these radars. (1) Accurate distance measurements to sharpen

up the whole study of planetary orbits. (2) Exploration of the planetary surfaces by studying the fine-structure of the radar echoes.

For test of general relativity we are concerned only with 1. But item 2 is so successful that I cannot resist discussing it first.

Range-Doppler Mapping. If a planet is rotating at the rate Ω radians per second, and has radius R, then a spot P on the planet



can be labelled with two coordinates x and r, apart from a North-South ambiguity which we will return to later (P and P' have the same values of x and r, where P' is the reflection of P in the equatorial plane). The line-of-sight velocity of P is then

$$v = V + \Omega x,$$

where V is the line-of-sight velocity of the center of mass of the planet.

It is convenient to take for the radial coordinate $y = (R^2 - r^2)^{1/2}$, so that the distance from the radar to a point (x,y) is

$$d = D - y + (V + \Omega x)t.$$

Suppose a transmitted signal has amplitude $a(t)$. Let $\rho(x,y)$ be the reflectivity at the point (x,y) on the planet. Then the reflected signal is

$$\psi(t) = \int dS\, \rho(x,y)\, a\left[t - \frac{2}{c}(D - y + (V + \Omega x)\, t)\right].$$

Or, taking the Fourier transform of this, where $\tilde{a}(\nu)$ is the transform of $a(t)$, $\tilde{\psi}(\nu)$ that of $\psi(t)$:

$$\tilde{\psi}(\nu) = \int dS\, \rho(x,y) \exp\left(\frac{2i\nu}{c}(D-y)\right) \tilde{a}\left[\nu + \frac{2\nu}{c}(V+\Omega x)\right].$$

Suppose we transmit a train of identical pulses with time-interval τ, each of them being in a frequency band around the radar frequency ω. Then $\tilde{a}(\nu)$ is a discrete spectrum consisting of δ-functions at the points $\nu = \omega + 2\pi k/\tau$, k integer.

Picture of Range-Doppler signal

In time

Transmitted signal $a(t)$

Reflected signal $\psi(t)$

Pulse spacing τ, coherent processing time T. Effective number of pulses T/τ equals number of resolution elements in *frequency*.

In frequency

Transmitted signal $\tilde{a}(\nu)$

Reflected signal $\tilde{\psi}(\nu)$

Line spacing $2\pi/\tau$, total band-width $2\pi/T'$, where T' is a rise-time of a single pulse. Effective number of lines τ/T' equals number of resolution elements in *range*. In the reflected signal each discrete line is spread from $x = +R$ to $x = -R$, i.e., between

$$\omega + 2\pi k/\tau + (2\omega/c)\,[V \pm \Omega R].$$

If these reflected bands do not overlap, then we can isolate the reflection from a particular value of x without ambiguity. The condition for nonoverlapping is

$$2\pi/\tau > 4\omega\Omega R/c.$$

Having fixed x, the observed $\tilde{\psi}(\nu)$ provide Fourier coefficients in a Fourier series expansion of $\rho(x,y)$ as a function of y. This Fourier series determines $\rho(x,y)$ unambiguously if and only if the pulse repetition time is long enough to avoid overlapping of the echoes from successive pulses, i.e.,

$$\tau > 2R/c.$$

We see that a suitable τ can be chosen if and only if

$$\Omega\omega R^2/c^2 < \pi/4.$$

This condition favors very low rates of rotation. For example, for Venus, Period = 250 days, $\Omega = 2.5 \times 10^{-7}$, $R^2/c^2 = 5 \times 10^{-4}$,

$$f = \omega/2\pi < c^2/8\Omega R^2 = 10^9 = 1000\,\text{Mc}.$$

This is just fine for Arecibo, which has $f = 430$ Mc. When you come to a rapidly rotating planet like Mars, things are much more difficult. Note that the actual speed of rotation at the Venus equator is

$$R\Omega = 4 \text{ miles per hour},$$

a brisk walking speed. It is velocities like this that the radars (with modern data processing) can handle best.

Let us just look at the resolution with which Range-Doppler mapping may be done.

Resolution in x. Since the reflected frequency is $\nu + 2\omega\Omega x/c$, we have

$$\Delta x = c\Delta\nu/2\omega\Omega = c/2\omega\Omega T,$$

where T is the coherent integration time of the data processing.

The pulse-repetition time must satisfy

$$\tau \leqq \pi c/2\omega\Omega R$$

so that

$$\Delta x \geqq \tau R/\pi T.$$

Resolution in y. This is just

$$\Delta y = cT'$$

where T' is the rise-time of the pulse. Since we choose

$$\tau > 2R/c,$$

we have

$$\Delta y > 2RT'/\tau.$$

The total number of resolution cells in (x, y) is

$$N = 2R^2/\Delta x \Delta y \leqq \pi T/T',$$

which shows that we just fulfill the uncertainty principle for simultaneous measurement of time and frequency.

Take the example of Arecibo observing Venus with 430 Mc radar. In this case

$$T = 50\,\text{sec},$$

$$\tau = 5 \times 10^{-2}\,\text{sec},$$

$$T' = 10^{-5}\,\text{sec}.$$

Then

$$\Delta x = c/2\omega\Omega T = 5\,\text{km},$$

$$\Delta y = cT' = 3\,\text{km}.$$

In practice we are not able to do as well as this. But we shall in time.

The beauty of the Range-Doppler method is that you have mapping with resolution *independent of distance,* provided you have enough power to get echoes from the planet at all. You are not dependent (except for resolution of the North-South ambiguity) on the angular resolution of the radar.

On the Moon it has actually been done with resolution of ~ 10 km, using the libration of the Moon to produce the Doppler effect.

Because of the North-South ambiguity, the (x, y) map gives not a true map of the radar reflectivity of the planet, but a superposition of the maps of the Northern and Southern hemispheres. In principle this ambiguity can be resolved by doing the entire mapping many times, when the Earth is oriented at various angles as seen from the planet.

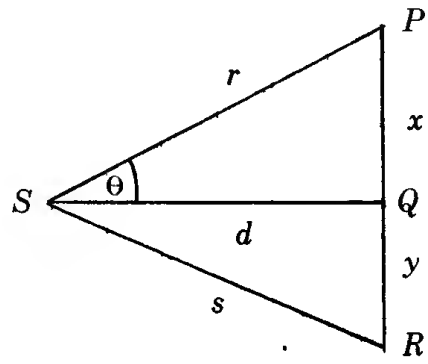
In practice the North-South ambiguity has been resolved only in the case of the Moon. For the Moon it is possible to use radar beams which are narrow enough to discriminate against one or the other hemisphere. For the planets this will presumably never be possible.

III. **The relativity experiment.**[1] The Sun produces a Schwarzschild metric

$$ds^2 = [1 - 2m/r]dt^2 - [1 - 2m/r]^{-1}dr^2 - r^2d\Omega^2$$

with $m = GM/c^2 = 1.5\,\text{km}$.

On a light ray $ds = 0$.



In moving from the point P to a neighboring point on the line PQ we have

$$r = d\sec\theta,$$

$$dr = d\sec\theta\tan\theta\,d\theta, \qquad d\theta = d\Omega;$$

hence

$$[1 - 2m/r]dt^2 = r^2d\theta^2 + r^2\tan^2\theta\,[1 - 2m/r]^{-1}d\theta^2,$$

and therefore

$$dt/d\theta = d\sec^2\theta + m\,[2\sec\theta - \cos\theta] \text{ to order } m/d,$$

$$t = d\tan\theta + m\,[2\log\tan(\theta/2 + \pi/4) - \sin\theta]$$

$$= x + m\,[2\log d/(r - x) - x/r].$$

Therefore on a straight path from P to R and back the elapsed time is

$$t = 2(x + y) + 4m\log[(r + x)(s + y)/d^2] - 2m\,[x/r + y/s].$$

[1] Professor Schiff in his article in these Lectures has already covered the material in the beginning half of this section. So we are discussing this in as brief a form as possible.

Now this time t is the time measured by a stationary clock at a great distance from the sun. The time measured on a circular orbit through P is given by

$$dt_p = [1 - 3m/r]^{1/2} dt = [1 - 3m/2r]dt.$$

Hence the elapsed time in the round trip $P \rightarrow R \rightarrow P$ measured by a clock carried by P is

$$t_p = t[1 - 3m/2r]$$
$$= 2(x + y) + 4m \log[(r + x)(s + y)/d^2] - m[(5x + 3y)/r + 2y/s]$$

to order m/d.

We have calculated t_p for a track which is a straight line in coordinate space. But this is equal to the elapsed time along the real light-ray, for the following reason.

The real light-ray satisfies the Fermat principle of least time

$$\delta t = 0$$

for small variations in the path from the physical path. Now the physical path deviates from the straight line by the angle

$$\Delta = 4m/d.$$

Thus $t_{\text{Physical}} - t_{\text{Straight}}$ is of order $\Delta^2 = 16m^2/d^2$ and is negligible to order m/d.

Now the question is, how do these coordinate distances x, y, r, s relate to "real" distances? What do x and y really mean? The answer is that x and y are fictitious quantities which must be calculated from a relativistic solution of the planetary orbits. We know that in terms of the Schwarzschild coordinates (r, θ) the planetary orbit is determined by

$$d^2(1/r)/d\theta^2 + 1/r = K + 3m/r^2,$$

which is exactly the Newtonian equation plus the "advance in perihelion" correction $3m/r^2$. So to determine x and y accurately to terms in m/r we would have to do a complete fitting of the orbits including relativistic corrections. The radar experiment is in principle not to be disentangled from a measurement of the perihelion advance and other relativistic effects.

However we can make a separation between terms of order m/d

and m/r if $d \ll r$, i.e., if the planet is almost behind the Sun as seen from the Earth.

If we are looking at Mercury at $s = 6 \times 10^7$ km from the Earth at $r = 15 \times 10^7$ km, and if Mercury is around $d = 4R_s = 3 \times 10^6$ km, i.e., about 1° in angle from the Sun, then

$$m/d \approx 5 \times 10^{-7}, \qquad m/r \approx 10^{-8}.$$

So we can consider m/r to be negligible while keeping terms in m/d. In this case x and y are simply the distances of the planets as determined by fitting Newtonian orbits to the observed motions, and

$$t_p = 2(x + y) + 4m \log(4rs/d^2) - m\,(7 + 3s/r).$$

At $d = 3 \times 10^6$ km, $\log(4rs/d^2) = \log 4000 = 8.3$;

Time delay $= 4m\,(8.3) - m\,(8.2) = 25.0m = 37.5\,\text{km}/c = 125\,\mu\text{sec}.$

Present-day timing accuracy would allow measurement to about $10\,\mu$ sec. Since the light-travel time is

$$2(r + s) = 1400\,\text{sec},$$

this is a measurement to 1 part in 10^8. So the delay could be measured to 10% accuracy. This is somewhat better than present-day measurements of the light-deflection. No doubt both will be improved in the future.

An objection can be made to the claim that this radar experiment is a "new" test of general relativity on the grounds that it measures only the *linear* terms in m/r in the Schwarzschild metric, which are already determined by the light-deflection and red-shift tests. (The perihelion advance also measures a term in $(m/r)^2$.) In fact, if you imagine that the space around the Sun has a refractive index $[1 - 2m/r]^{-1}$ you find the correct light-deflection, and this combined with the red-shift formula gives the correct radar time-delay formula.

However this interdependence of the three tests is true only if you suppose a metric exists, i.e., you assume something like general relativity to start with. If one wishes to test G.R. one should imagine what might happen if G.R. were false. For example if light consisted of Newtonian particles with special-relativistic mechanics, the light-deflection would be $+1/2$ the Einstein value, while the

time-delay would be zero. Actually Newton would have assumed nonrelativistic mechanics for his photons, which would give

$$t_{\text{Newton}} = 2(x+y) - 2m \log [(r+x)(s+y)/d^2],$$

a delay of $(-1/2)$ times the Einstein value. (This remark is due to I. I. Shapiro.) So the radar experiment is looking directly at light-propagation in a gravitational field in a way which is in principle independent of the other tests.

Another version of the radar measurement is to look at the Doppler-shift produced by the time-delay, in addition to ordinary orbital Doppler-shifts which can be computed from the known orbits.

The additional Doppler-shift for a signal at frequency ω is

$$\delta\omega = -\omega \dot{t}_p = -\omega (dt_p/dd)\,[\dot{d}/c] = \omega (8m/d)\,[\dot{d}/c]$$

when $d \ll s, r$.

Now

$$\dot{d} = (rs/r+s)\,(\Omega_R - \Omega_P)$$
$$= 4 \times 10^7 [8 \times 10^{-7} - 2 \times 10^{-7}] = 24\,\text{km/sec} = 8 \times 10^{-5} c$$

for Mercury.

For Haystack

$$f = 8 \times 10^9 \text{ cycles},$$

$$\delta f = 5 \times 10^6\, m/d \text{ cycles} \approx 2 \text{ cycles when } d = 3 \times 10^6 \text{ km}.$$

Measurement of Doppler-shift to 2 cycles in 10^{10} is quite possible. But there are difficulties. Especially as Mercury is rotating at $\Omega \approx 2.5 \times 10^{-6}$, there is a Doppler broadening of

$$\delta\omega \approx R\Omega\omega/c = 2.5 \times 10^{-8}\omega, \qquad \delta f \sim 200 \text{ cycles},$$

which makes measurement to 2 cycles accuracy hard.

Experimental Difficulties. The greatest difficulty for both time-delay and Doppler measurements is the interplanetary plasma which has a high density near the Sun, and is not even constant in time. By corona spectroscopy, the electron density is measured to be (quiet sun conditions)

$$N = 5 \times 10^5 (R_s/r)^2 \text{ electron/cm}^3.$$

This gives a time delay

$$\Delta t_{\text{Plasma}} = (6.5 \times 10^{24}/f^2 d)\,[\arctan(x/d) + \arctan(y/d)]$$
$$= 2 \times 10^{25}/f^2 d \qquad \text{for } x, y \gg d.$$

Here f = frequency in cycles, d in cm.

So for Arecibo

$$f = 4.3 \times 10^8 \quad \text{and} \quad d = 3 \times 10^{11},$$

hence

$$\Delta t_{\text{Plasma}} = 4 \times 10^{-4}\,\text{sec.}$$

But for Haystack $f = 8 \times 10^9$, which gives

$$\Delta t_{\text{Plasma}} \cong 10^{-6}\,\text{sec.} \ll \Delta t_{\text{Relativity}}.$$

By measurement at several frequencies we can observe the plasma and relativity effects *separately*.

Other difficulties arise from the uncertainties in the *size* and *shape* of the planets themselves. These effects will become important when one tries to measure the path-lengths to 1 km accuracy.

There are proposals to place an artificial planet in orbit around the Sun, carrying a radar transponder. This would remove the problems connected with rotation Doppler broadening and planetary shape. But it is a difficult experiment, and one would have trouble computing drag effects of radiation and solar wind on the orbit of the probe. MIT plans to have such a probe in orbit in approximately 2 years.

IV. **Summary.** Summing up, the radar time-delay experiment gives a good check of general relativity, independent of other tests, if one works only to the presently attainable accuracy of about 10% of the delay (accuracy of 1 part in 10^8 in the time measurement).

When the measurement can be done to significantly higher accuracy (say 1% or better), then it is no longer independent of other relativistic effects. So in the long run one must think of the radar measurements as a single large program in which all of the effects

Planetary orbits and perturbations,
Perihelion precession,
Gravitational time-delay,
Interplanetary plasma,

Planetary sizes and shapes and rotations,
Planetary Range-Doppler maps

are simultaneously observed. The relativistic effects are finally to be extracted from a comparison of *all* the observations with a big machine-calculation of the planetary motions.

This big program has already begun. And in fact we already obtained from radar observations improvements in the planetary orbits which are roughly

a factor of 10 in the shapes,
a factor of 1000 in the absolute scale

compared with the best that optical observations could do.

In principle, optical astronomy is accurate to about 10^{-7} in angle (0.01 second of arc) and gives very poor information on absolute scale. Radar is accurate to about 10^{-8} in relative distances, and to about 10^{-6} in absolute distance (because c is still uncertain to 1 part in 10^6).

Thus radar observations in a few years will give improved accuracy in the most sensitive test of general relativity, the perihelion motion of Mercury.

The relativistic precession is about 43″ per century which means an actual displacement of Mercury by about

$$6 \times 10^7 \times 2 \times 10^{-5} \times 1/5 \text{ km per year} = 24 \text{ km/year}$$

(1/5 because the orbit is not very eccentric). If we can measure distances to 1 km then in 10 years we have the perihelion advance to an accuracy of 4×10^{-3}. In fact in 10 years the accuracy of radar will no doubt be considerably better than this.

[*Note added for second edition.*] Since this chapter was written a book "Radar Astronomy" by J. V. Evans and T. Hagfors (McGraw-Hill, New York, 1968) has appeared, giving a good general survey of the subject. The most recent (1971) results of the radar measurements of planetary orbits are reported by I. I. Shapiro and colleagues in Phys. Rev. Lett. **26** (1971), 27 and 1132. The progress in the period 1967-1971 has been a little slower than expected because of shortage of funds both for observing and for computation. But the relativistic time-delay of radar signals passing the sun is now reliably measured to better than 10% accuracy, and agrees with the Einstein prediction.

INSTITUTE FOR ADVANCED STUDY
PRINCETON, NEW JERSEY

R. K. Sachs

Gravitational Waves[1]

Accelerating masses radiate gravitational waves much as accelerating charged particles radiate electromagnetic waves. The motivation for studying gravitational waves is (i) to understand the fundamental physics, (ii) to master the seductive mathematics, and (iii) to calculate numbers which might be measurable.

Rather than presenting a general introduction to gravitational radiation I will present here several detailed examples. I shall first discuss exact solutions, then approximation methods.

I. **Exact solutions—Plane waves.**

1. As an introductory example consider a plane electromagnetic wave in Minkowski space. Suppose the wave is traveling in the z direction with velocity $c = 1$.

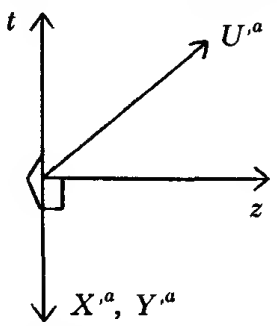
Define the following quantities:

$$\eta_{ab} = \begin{pmatrix} -1 & & & 0 \\ & -1 & & \\ & & -1 & \\ 0 & & & 1 \end{pmatrix},$$

[1] Research supported by Aerospace Research Labs., OAR, AF-33 (615)-1029.

$$U_{,a} = (0,0,-1,1) \Rightarrow U^{,a} = (0,0,1,1),$$
$$X_{,a} = (1,0,0,0) \Rightarrow X^{,a} = (-1,0,0,0),$$
$$Y_{,a} = (0,1,0,0) \Rightarrow Y^{,a} = (0,-1,0,0)$$
$$\Rightarrow U^{,a}U_{,a} = U^{,a}X_{,a} = U^{,a}Y_{,a} = 0.$$

The wave fronts are given by $U = t - z =$ constant, the propagation vector is $U^{,a}$, and the two possible polarization vectors are



$X^{,a}$ and $Y^{,a}$. I will just write down the electromagnetic tensor F_{ab} and show that it satisfies Maxwell's equations for a vacuum,

$$F_{ab} = -2G(U)\,X_{,[a}U_{,b]} - 2H(U)\,Y_{,[a}U_{,b]},$$

where $G(U)$ and $H(U)$ are arbitrary functions of the phase U. Computing $F^{ab}{}_{,b}$ we have

$$F^{ab}{}_{,b} = -2G'(U)\,U_{,b}X^{,[a}U^{,b]} - 2H'(U)\,U_{,b}Y^{,[a}U^{,b]}.$$

Since $U_{,b}U^{,b}$, $U_{,b}X^{,b}$ and $U_{,b}Y^{,b} = 0$ we have $F^{ab}{}_{,b} = 0$ and similarly $F_{[ab,c]} = 0$, which are the Maxwell equations for vacuum. To see what the field is like we take $H(U) = 0$ and compute $E_x = F_{14}$ and get $E_x = G(U)$; similarly $H_y = E_x$. For monochromatic plane waves we take $G(U) = e^{i\omega U} = e^{i\omega(t-z)}$. The other polarization ($G(U) = 0$) would give $E_y = -H_x = H(U)$.

A possible potential is $A_a = U_{,a}[2Gx + 2Hy]$, which gives $F_{ab} = A_{[b,a]}$.

2. *Gravitational plane waves*, $G^{ab} = 0$. I will write down a form for the metric, then say briefly where it came from, and then check the solution in some detail.

The metric is [1]

$$g_{ab} = \eta_{ab} + U_{,a}U_{,b}A, \quad A = G(U)(x^2 - y^2) + 2H(U)\,xy,$$

where $U = t - z$ as before so that

$$ds^2 = -dx^2 - dy^2 - dz^2 + dt^2 + (dt - dz)^2 A.$$

One might ask: Where did this solution come from? Well, the quantities η_{ab} and F_{ab} in the electromagnetic case are invariant under a 5 parameter subgroup of the Lorentz group, namely, translations in the x, y, and $(0,0,1,1)$ directions plus two rotations in the $U^{,a}, X^{,b}$ and $U^{,a}, Y^{,b}$ planes. Assuming an identical symmetry group [2] in the gravitational case where $G_{ab} = 0$ we find there exist coordinates x, y, z, t in which the metric has the above form.

Assuming the above form for the metric we want G_{ab} and get it by just computing g^{ab}, then Γ^a_{bc}, then R_{abcd}, and then R_{ab}.

If we define $U^{,a} = \eta^{ab} U_{,b}$ and $g^{ab} = \eta^{ab} - U^{,a} U^{,b} A$ it follows that

$$\begin{aligned} g^{ab} g_{bc} &= (\eta^{ab} - U^{,a} U^{,b} A)(\eta_{bc} + U_{,c} U_{,b} A) \\ &= \delta^a_{\ c} - U^{,a} U_{,c} A + U^{,a} U_{,c} A = \delta^a_{\ c}, \end{aligned}$$

and at the same time one checks:

$$\begin{aligned} &U^{,a} = g^{ab} U_{,b}, \\ &X_{,a} = (1,0,0,0) \Rightarrow X^{,a} = (-1,0,0,0), \\ &Y_{,a} = (0,1,0,0) \Rightarrow Y^{,a} = (0,-1,0,0), \\ &U^{,a} U_{,a} = U^{,a} X_{,a} = U^{,a} Y_{,a} = 0, \text{ as before.} \end{aligned}$$

We can calculate the Christoffel symbols and the Riemann tensor:

$$\Gamma^a_{bc} = \frac{1}{2} g^{ae} \{ g_{eb,c} + g_{ec,b} - g_{cb,e} \},$$

$$R_{abcd} = -g_{ae} \Gamma^e_{bc,d} + g_{ae} \Gamma^e_{db,c} - g_{ae} \Gamma^e_{d\lambda} \Gamma^\lambda_{cb} + g_{ae} \Gamma^e_{c\lambda} \Gamma_{db},$$

where Schild's conventions are used here and throughout.

One obtains

$$\begin{aligned} R_{abcd} = -4 \{ & G(U) [U_{,[a} X_{,b]} U_{,[c} X_{,d]} - U_{,[a} Y_{,b]} U_{,[c} Y_{,d]}] \\ & + H(U) [U_{,[a} X_{,b]} U_{,[c} Y_{,d]} + U_{,[a} Y_{,b]} U_{,[c} X_{,d]}] \}. \end{aligned}$$

$G(U)$ and $H(U)$ will presently be interpreted as the two amplitudes in the two possible polarization modes. The propagation vector is again $U^{,a}$. Computing $R_{bc} = g^{ad} R_{abcd}$ we get

$$R_{bc} = -4G(U) \left[\frac{1}{4} U_{,[b} U_{,c]} - \frac{1}{4} U_{,[b} U_{,c]} \right] + 0 = 0,$$

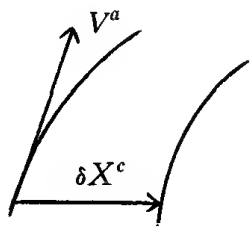
and therefore the Einstein tensor vanishes and we have a vacuum solution.

The Petrov type can be read off from the equation for R_{abcd} by noticing $R_{abcd}U^{,d} = 0$. The type is therefore N. (See Schild's lectures.)

3. *Interaction of gravitational waves with test particles.* Now imagine a gravitational wave impinging on two test particles. Actual test particles may be two atoms in a lattice, two weights on a spring, or any other physical object whose characteristic frequency corresponds to that of the incoming gravitational waves. The frequencies should be matched to assure the maximum energy absorption. In practice the dimensions of the test device are then very small compared to the wave length. Moreover, if we know relative accelerations in absence of atomic forces the actual relative accelerations can be computed. Therefore we replace the actual test device by two freely falling bodies and apply the equations of geodesic deviation:

$$\delta\ddot{X}^a = R^a{}_{bcd}\, V^b V^c \delta X^d,$$

where V^a is the world velocity of one of the two test particles and δX^a the connecting vector.



$$V^a V_a = 1, \quad \delta X^a V_a = 0.$$

Assume $V^b = (0, 0, \alpha, \beta)$, so that V^a lies in the z, t plane; this specialization does not affect the final conclusions. For the term $E^a{}_d \equiv R^a{}_{bcd}\, V^b V^c$ we have

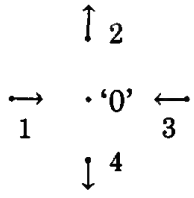
$$E^a{}_d = (V^b U_{,b})^2 \{ G(U) [X^{,a} X_{,d} - Y^{,a} Y_{,d}] + H(U) [X^{,a} Y_{,d} + Y^{,a} X_{,d}] \}.$$

$dU/ds = U_{,b} V^b$ is the rate of change of phase seen by an observer with velocity V^b. It can also be interpreted as a Doppler correction term to the frequency. To see this consider the case $G(U) = e^{i\omega_0(t-z)}$. Then

$$\frac{dG(U)}{ds} = G'(U)\frac{dU}{ds} = G(U)\,[i\omega_0 U_{,b} V^b],$$

where s is the proper time of the observer. In the term $E^a{}_b$ above, $(V^b U_{,b})^2$ can be interpreted as a Doppler shift correction to the wave amplitudes $G(U)$ and $H(U)$.

For $H(U) = 0$ consider a wave incident normally on the paper which is taken as the x, y plane. Suppose '0' is an observer at $x = y = 0$ in the plane, and he is observing 6 other neighboring particles. One is above and one below '0' out of the plane of the paper, and the other 4 are as shown below. The two particles out of the plane of the paper undergo no acceleration relative to '0', and the other 4 particles accelerate only in the plane as indicated below.



The $G(U) = 0$ polarization is just rotated by 45°.

If these were velocities rather than accelerations the fields would be expansion and rotation free, but of course not shear free. The wave therefore excites quadrupole shear oscillations of the detector.

The metric $ds^2 = -dx^2 - dy^2 - dz^2 + dt^2 + (dz - dt)^2 A$ appears to have a singularity as x or $y \to \infty$, but it does not. There exist coordinates which exhibit this lack of singularity, but the metric takes on a more complicated form. Alternatively, there exist Killing vectors which take you from any one point in the x, y, U hypersurface to any other point, and the nonsingularity at one point insures the nonsingularity at any other [1], [3]. Moreover, the space is geodesically complete [3]. That is, along all geodesics, spacelike, lightlike, or timelike, you can go to arbitrarily large values of the affine parameter.

4. *Other exact solutions.* A solution of the field equations $G_{ab} = -8\pi K T_{ab}$ with nonvanishing T_{ab} which contains a gravitational wave from a bounded, reasonable source, so far has not been found. What one can try to do is solve the equations $G_{ab} = 0$ outside the matter and then try to deduce some properties of the source by

looking at the singularities in the solution. There are two general methods of solving the equations $G_{ab} = 0$. They are,

(1) Group theoretical methods [2],

(2) Petrov-Pirani theory; Ehlers, [4].

In the end these methods do not give numbers for physical situations of interest. (1) breaks down because of the lack of symmetry in a radiation field from a bounded source, and (2) breaks down because solutions that are algebraically special, in some sense, do not contain self-scattering of the waves, and hence they can come only from rather weird sources.

II. **Approximation methods.** Because the exact solutions discussed in the first lecture do not show how gravitational radiation is related to its sources they do not produce useful numbers. To get estimates of the amount of radiation emitted by, say, an exploding supernova, we must turn to approximation methods. One of the oldest approximation methods, and still the most useful, is to linearize the field equations. Most of this lecture will be concerned with linearized theory and the numerical estimates that it leads to. However, we shall also need to discuss briefly two other methods: asymptotic approximations and slow-motion approximations.

1. *Linearized approximation.* Suppose there exist coordinates in which the metric has the form

$$g_{ab} = \eta_{ab} + \epsilon h_{ab} \tag{1}$$

where ϵh_{ab} is in some sense small. We can then systematically expand all relations of interest in powers of ϵ and drop all but the first nontrivial term in each equation; this procedure gives linearized theory. ϵ is treated as a formal expansion parameter which is set equal to one at the end of the calculation.

If one plays with linearized theory awhile one finds some rules of the game. Before going into formal details let me mention these rules explicitly because on occasion different people mean slightly different things by linearized theory. (a) In units $c = G = 1$, which we shall use throughout, let M be a typical source mass, r a distance from the source, L a typical source dimension, and ω a typical source frequency. Then the order of magnitude of the dimensionless quantity ϵh_{ab} is typically, in the far zone,

$$\text{(2)} \qquad \text{(a)} \quad M/r \qquad \text{or} \qquad \text{(b)} \quad ML^2\omega^2/r$$

in static and dynamic situations respectively. Since $L\omega \cong V < 1$ where V is a typical velocity with which one part of the source moves relative to other parts, (2b) is typically smaller than (2a). (b) The sources of a linearized field live in Minkowski space; they don't know anything about gravitational fields.[2] (c) Of course, linearized gravitational fields do not scatter off themselves. (d) In case of doubt it always pays to regard linearized theory as an approximation to the exact theory rather than as a theory in its own right. However, after awhile one gets bored with this attitude because it almost always emerges that h_{ab} can be considered as a Lorentz covariant field subject to Lorentz covariant field equations; all that remains of the general coordinate transformations are the gauge transformations discussed below. Because of this fact we shall henceforth raise and lower all indices with η_{ab} and η^{ab} respectively. (e) The effect of a linearized field on a test device or test photon can be obtained just as in the exact theory—the field may act on detectors though it is not allowed to react on sources.

2. *Gauge transformations.* Aside from Lorentz transformations, there are other coordinate transformations that preserve the form (1) with ϵh_{ab} small, namely

$$x'^a = x^a + \epsilon\xi^a(x) \tag{3}$$

with $\xi^a(x)$ arbitrary. From

$$g_{ab} = \frac{\partial x'^c}{\partial x^a}\frac{\partial x'^d}{\partial x^b} g'_{cd}$$

and (1) we have

$$h'_{ab} = h_{ab} - 2\xi_{(a,b)} \quad \text{where} \quad \xi_{(a,b)} \equiv 1/2(\xi_{a,b} + \xi_{b,a}). \tag{4}$$

The transformations (3) and (4) are called gauge transformations in analogy to the transformations $A'_a = A_a - \phi_{,a}$ of electrodynamics. Incidentally (4) is a special example of the following general theorem. Suppose $g_{ab} = g_{0ab} + \epsilon h_{ab}$ with arbitrary g_{0ab} and $T_a{}^{b\cdots}{}_{\cdots}$ is any tensor function of g_{ab} with expansion $T_a{}^{b\cdots}{}_{\cdots} = T_{0a}{}^{b\cdots}{}_{\cdots} + \epsilon T_{1a}{}^{b\cdots}{}_{\cdots} + \cdots$. Then under a transformation (3)

[2] In particular we temporarily exclude from the discussion sources such as rotating double stars whose acceleration is caused by gravitational fields.

$$T'_{1a\ldots}{}^{b\cdots} = T_{1a\ldots}{}^{b\cdots} + \mathscr{L}_\xi T_{0a\ldots}{}^{b\cdots} \tag{5}$$

where $\mathscr{L}_\xi$ is the Lie derivative (see Schild's paper in these Lectures). Only gauge invariant quantities really mean anything in linearized theory because the gauge transformations are essentially just coordinate transformations.

2. *Field equations and their solutions.* Equation (1) and $g_{ab}g^{bc} = \delta_a^c$ implies $g^{ab} = \eta^{ab} - \epsilon h^{ab}$ to first order in ϵ. Also,

$$R_{abcd} = -\epsilon(h_{a[c,d]b} - h_{b[c,d]a}) + O(\epsilon^2), \tag{6}$$

$$G_b^a = \epsilon/2(\gamma^{ab,c}{}_c - \gamma^{ac,b}{}_c - \gamma^{bc,a}{}_c + \eta^{ab}\gamma^{cd}{}_{,cd}), \tag{7}$$

where γ^{ab} is defined by $\gamma^{ab} = h^{ab} - 1/2\eta^{ab}h^c_c$. Henceforth we shall omit the symbols $O(\epsilon^2)$, etc. It follows from (6), (7), and (4), or directly from (5) that R_{abcd} and G_b^a are gauge invariant. It follows from (7) or from expanding $G^{ab}{}_{;b} \equiv 0$ that

$$G^a_{b,a} = 0. \tag{8}$$

In units $c = G = 1$ the field equations read

$$G^{ab} = -8\pi T^{ab} \tag{9}$$

and are consistent with (8) since

$$T^{ab}{}_{,b} = 0 \tag{10}$$

for Lorentz covariant sources. (9) must in general be supplemented by some equations which govern the time development of the sources, but in linearized theory one can regard the sources as given. (9) has as one solution the retarded integral

$$\gamma^{ab} = -4\int \frac{(T^{ab})_{\text{ret}}}{|\mathbf{x} - \mathbf{x}'|}\, d^3x' \tag{11}$$

where we now set $\epsilon = 1$. This solution can be guessed from analogy with electrodynamics. It can be verified directly by substituting into (7), noting that

$$\begin{aligned} \Box\left\{\frac{f(\mathbf{x}', t')_{\text{ret}}}{|\mathbf{x} - \mathbf{x}'|}\right\} &= -4\pi f(\mathbf{x}', t')_{\text{ret}}\,\delta^3(\mathbf{x} - \mathbf{x}') \\ &= -4\pi f(\mathbf{x}, t)\,\delta^3(\mathbf{x} - \mathbf{x}') \end{aligned} \tag{12}$$

where $\delta^3(x)$ is the three-dimensional delta function and $\Box$ the D'Alembertian operator. Note that $T^{ab}{}_{,b} = 0$ together with (11)

implies

$$\gamma^{ab}{}_{,b} = 0. \tag{13}$$

For reasonable T^{ab} one can show the following uniqueness theorem: The general solution of (9) that has no incoming radiation differs from (11) only by a gauge transformation [4].

Suppose that the center of mass frame for the Lorentz covariant quantity T^{ab} is well defined and that $T^{ab} = 0$ in that frame whenever $r = (x^2 + y^2 + z^2)^{1/2} > r_0$. Then one can show that for calculating γ_{ab} (or R_{abcd}) at points $r > r_0$, T^{ab} in (11) can be replaced by a multipole series of the form

$$T^{ab} = A^{ab}(t)\,\delta^3(\mathbf{x}) + B^{ab\alpha}(t)\,\delta^3(\mathbf{x})_{,\alpha} + \cdots \qquad (\alpha = 1,2,3), \tag{14}$$

just as the charge current four-vector j^a can be expanded in multipoles in electrodynamics. We shall analyze the resulting fields up through quadrupole terms.

If one requires that a T^{ab} of the form (14) obey (10), that T^{ab} contain no terms higher than quadrupole terms, and that we work in the center of mass frame, then one gets by a slightly tedious calculation omitted here [5]:

$$T^{ab} = m\,V^aV^b\delta^3 + 1/2(L^{ac}V^b + L^{bc}V^a)\,\delta^3{}_{,c} + 1/2(Q^{aibj}(t)\,\delta^3)_{,ij} \tag{15}$$

where

$$V^a = \delta^a_4, \quad m = \text{const}, \quad L^{ac} = -L^{ca} = \text{const}, \quad L^{ab}V_b = 0 \tag{16}$$

and

$$Q^{aibj} = Q^{[bj][ai]}, \quad Q^{a[ibj]} = 0, \quad Q^{aibj}\eta_{ij} = 0. \tag{17}$$

In (15) certain terms which give rise to no field outside the source or to a field which can be eliminated by a gauge transformation have been omitted. Note that Q^{aibj} has all the symmetries of a conformal tensor and therefore has 10 independent components. What do the various terms mean? Well, from integrating (15) and using (16) and $\int \delta^3{}_{,\alpha}\,d^3x = \int \delta^3{}_{,\alpha\beta}\,d^3x = 0$ $(\alpha,\beta = 1,2,3)$ we have

$$m = \int T^{44}d^3x. \tag{18}$$

So m is the mass or monopole moment. Similarly (15), (16) and $\int \delta^3{}_{,\alpha}\,x^\beta d^3x = -\delta^\beta_\alpha$ etc. imply $L^{a4} = 0$ and

$$L^{\alpha\beta} = -\int (T^{4\alpha}x^{\beta} - T^{4\beta}x^{\alpha})\, d^3x. \tag{19}$$

So $L^{\alpha\beta}$ is essentially the angular momentum. Next we have by similar manipulations

$$Q^{4\alpha 4\beta}(t) = \int T^{44}(x^{\alpha}x^{\beta} - (1/3)\,\delta^{\alpha\beta}r^2)\, d^3x = \int T^{44}x^{\alpha}x^{\beta}d^3x. \tag{20}$$

Therefore $Q^{4\alpha 4\beta}(t)$ represents vibrational quadrupole moments. The remaining components of Q^{aibj} represent rotational quadrupole moments; we shall not analyze these further here, because their contribution is usually small compared to that of the vibrational quadrupole moments.

The solution of $G^{ab} = -8\pi T^{ab}$ with T^{ab} of the form (15) is simply

$$\begin{aligned}\gamma^{ab} &= m V^a V^b/r + 1/2(L^{ac}V^b + L^{bc}V^a)(1/r)_{,c} \\ &\quad + 1/2(Q^{aibj}(U)/r)_{,ij}\end{aligned} \tag{21}$$

where $U = t - r$ is the retarded time. This solution can be obtained by integrating (11); to check that it is a solution we merely use (7), (12) and its derivatives, and note that $\gamma^{ab}{}_{,b} = 0$ is implied by (16), (17), and (21).

From (6) and (21) we can now get R^{abcd}. Let us retain only the far-field, $1/r$ terms. Then using (17)

$$R^{abcd} = (-2/r)(\dddot{Q}^{iaj[c}U^{,d]}U^{,b} - \ddddot{Q}^{ibj[c}U^{,d]}U^{,a})\, U_{,i}\, U_{,j}, \tag{22}$$

where $Q^{aibj}(U) \equiv dQ^{aibj}/dU$ etc. Note various features of this result: (i) The monopole and dipole moments do not contribute to the radiation field; this result is analogous to the result in classical electrodynamics that monopoles do not radiate. (ii) $R^{abcd}U_{,d} = O$ to this order and $U_{,a}U^{,a} = O$ so the far-field is Petrov type N; this result implies that in any one direction the polarization pattern is the same as that I sketched in §I for plane waves. (iii) Apart from small numerics, polarization directions and angular variations, (22) means

$$\delta\ddot{x} \cong (\ddddot{Q}/r)\,\delta x \tag{23}$$

where Q is some source quadrupole moment and δx the separation between two test-particles as in §I.

Equation (23), or more explicitly Equation (22), together with (20) are the key equations for establishing numerical estimates.

Suppose we insert some numbers appropriate to astrophysics. We have a body of roughly solar mass $M_\odot \cong 10^{-5}$ sec and solar size $L_\odot \cong 10$ sec which has a quadrupole moment of order $M_\odot L_\odot^2$ from (20) provided it is quite irregular. It undergoes some violent explosion. Then a typical frequency ω of the motion cannot be larger than $1/L_\odot$ because otherwise the velocities involved would be greater than $c = 1$. Finally suppose our distance from the event is 1000 psc $\cong 10^{11}$ sec. Then the radiation from (23), causes

$$\ddddot{Q}/r \cong (M_\odot L^2{}_\odot \omega^4)/r < 10^{-18}\,\mathrm{sec}^{-2}. \tag{24}$$

For comparison, a peanut at rest in college town about a mile away causes static tidal forces—not forces, but tidal forces since we are talking of relative accelerations per unit distance—of roughly the same magnitude. Of course a time dependent signal might be easier to detect than a static peanut, but the numbers are clearly bad.

One can try inserting other situations from H-bombs to quasars into (23) but nothing really buys you much unless one lets L get nearly as small as $2M$ so that ω can in principle get nearly as large as $1/2M$. Of course for $L < 2M$ the body has gone through its own Schwarzschild "singularity" and all effects are red shifted beyond detectability. Thus for an exploding neutron star under the most extreme conditions $1/\omega \lesssim L \lesssim 2M_\odot$ we would have

$$\delta\ddot{x}/\delta x \cong 1/(4M_o r) \lesssim 2 \times 10^{-7}\,\mathrm{sec}^{-2} \tag{25}$$

provided we can trust linearized theory in this now heavily non-linear region. This might barely be detectable. Note that considering heavier bodies only decreases the relative accelerations caused by the gravitational waves (though, as Misner pointed out during the lecture, it increases the length of time the signal could last) and of course known lighter bodies have dimensions much larger than their own mass. Therefore (25) is close to an absolute upper bound in nature as we know it.

I should like to conclude with two comments. First, what about radiation from a pair of stars (or galaxies) that circle each other due to their Newtonian gravitational attraction [5]. This problem is different from those considered above for two related reasons:

(1) The force that here causes accelerations is itself a gravitational field, so that our picture of a T^{ab} caused by sources which do not notice gravity breaks down.

(2) The velocities involved automatically obey $V^2 \cong M/r$; therefore linearized theory which neglects terms of order $(M/r)^2$ but keeps terms of all orders in V does not make much sense. To really handle the problem one has to go to the so-called slow motion approximation in which everything gets expanded in powers of V (remember $V < c = 1$). Then life gets very complicated because the radiation effects, if they show up at all, will show up only in quite high order terms. Until very recently it was not clear whether one got radiation or not. However, it now seems to be agreed that in fact the Equations (20), (22) from linearized theory also apply to that case [6].

The second comment concerns radiation reaction. Presumably as the source sends out waves it in some sense loses energy and has its motion damped. No one has ever really managed to work out this effect in convincing detail. However, calculations in post-linear approximation using the pseudo-tensor [5] and some more rigorous asymptotic (large distance) approximations [7] both suggest the following: (1) The source does lose energy; (2) The amount of energy loss is that obtained from the pseudo-tensor calculation. The latter calculation shows that the power loss (again suppressing numerics and polarizations) is the dimensionless quantity

$$(\dddot{Q})^2 \cong M^2 L^4 \omega^6. \tag{25}$$

Under extreme conditions $1/\omega \cong L \cong 2M$ this becomes a pure number of order unity—the source loses about one second of mass in one second of time $\cong 10^{59}$ ergs/sec. Then if we assume that even in this heavily nonlinear region the total energy available to the source is not larger than M it follows that the radiation under the most extreme conditions cannot last longer than a time of order M.

References

1. H. Bondi, F. Pirani and I. Robinson, Proc. Roy. Soc. London Ser. A 251 (1959), 519.

2. A. S. Petrov, *Einstein-Räume*, Akadamie-Verlag, Berlin, 1964.

3. L. Witten, *Gravitation*, Wiley, New York, 1962, Chapter 2.

4. J. Ehlers, *Lectures on gravitational waves, in: Relatività Generale*, Centro Internazionale Matematico Estivo, Rome, 1965.

5. L. Landau and E. Lifshitz, *The classical theory of fields*, Addison-Wesley, Reading, Mass., (1951) §11.2 W. Tulciew, Bull. Acad. Polon. Sci. Cl. III, 5 (1957), 279.

6. A. Trautman, Report to the London Conference on General Relativity, 1965.

7. R. Penrose, Phys. Rev. Lett. (2) **10** (1963), 66.

UNIVERSITY OF TEXAS
AUSTIN, TEXAS

J. Weber

Measurements of Space Time Curvature and Search for Gravitational Radiation

It has been said that roughly 5000 theoretical papers have been written about general relativity and less than a dozen experiments carried out. I will discuss a new kind of experiment which is being performed at the University of Maryland, using laboratory sized masses and the masses of the earth and moon.

A large fraction of published work concerns the tightly coupled problems of energy, radiation, and the equations of motion. Our experiments involve apparatus to measure the dynamical curvature tensor and search for gravitational radiation.

Sachs has already discussed the theory, but a few additional comments are in order now. To prove, theoretically and rigorously, that general relativity predicts gravitational radiation we must do the following:

(a) Obtain solutions of $R_{\mu\nu} = 0$ without singularities. These should resemble spherical waves at intermediate distances, but at large distances the boundary conditions should be appropriate to the metric of a Friedman type universe.

(b) In the region of the sources of the radiation a solution of

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \frac{8\pi G}{c^4} T_{\mu\nu}$$

must be obtained, again without singularities, and these solutions must be matched up at some appropriate boundary.

Various modifications of such a program in certain levels of approximation have been carried out. These investigations make the existence of radiation very plausible, but are not in any sense rigorous proofs. It is interesting that the corresponding mathematical problem for electrodynamics—the radio antenna—was not solved until 1939. By then it was an anticlimax.

Einstein gave us approximate, linearized solutions. He wrote ($\delta_{\mu\nu}$ is the Lorentz metric, $h_{\mu\nu}$ and $\phi_{\mu\nu}$ are first order quantities):

$$g_{\mu\nu} = \delta_{\mu\nu} + h_{\mu\nu}, \qquad \phi_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\,\delta_{\mu\nu} h,$$

$$\phi^{\mu\nu}{}_{,\nu} = 0, \qquad \Box\,\phi_{\mu\nu} = -\,16\pi \frac{G}{c^4} T_{\mu\nu}.$$

Hardly any physicist who looks at these equations doubts the existence of gravitational waves. It is interesting that Einstein changed his mind about this several times during his life. There are important questions such as the issue of the convergence of the series for which the linear approximation is the first term and the question whether there are exact solutions which indeed correspond to the linearized ones.

For the elementary particle physicist the linearized equations imply that gravitons are particles of spin 2. If we think in terms of a Yukawa potential, the infinite range of the gravitational field implies a zero rest mass for the gravitons. In a multipole expansion the lowest order radiative term is quadrupole radiation. This may be understood in simple terms as a consequence of momentum conservation. For example, consider an isolated system composed of a large mass M coupled to a small mass m by some means such as a spring. From linear momentum conservation we may write

$$M\dot{x}_M + m\dot{x}_m = 0.$$

For small velocities this equation may be differentiated to give, approximately,

$$M\ddot{x}_M + m\ddot{x}_m = 0.$$

This states that the sum of mass times acceleration over a system

vanishes, and this corresponds to the object which gives dipole radiation in electrodynamics. Retardation effects then come in to give quadrupole radiation as the lowest nonvanishing order.

Einstein and later Eddington considered radiation from a spinning rod. Their result for the radiated power was

$$P = 32\,GI^2\omega^6/5c^5.$$

Here G is the constant of gravitation, I is the moment of inertia, ω is the angular velocity, and c is the speed of light. For a one meter rod spun so rapidly that it is on the verge of rupture, the radiated power is 10^{-37} watts, and the radiative "lifetime" is 10^{35} years. Ordinary double stars have a radiative lifetime of about 10^{12} years.

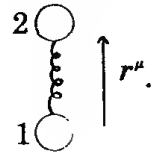
For atomic transitions, if the selection rules permit graviton emission, we can estimate the transition rate using the formulas from electrodynamics and replacing the square of the electron charge e by Gm^2. Since $Gm^2/e^2 = 10^{-43}$, this tells us that graviton emission is quite negligible in atomic processes.

These sources are hardly promising, but they suggest two kinds of experiments. One involves laboratory sources—the analogue of Hertz's experiments. In Einstein's day the power which could be detected was roughly sixty orders larger than the available power. Our work has improved this picture by about fifty orders. But a Hertz type of experiment today would perhaps involve a source of the size of a small campus filled with stress and energy to the bursting point, and suitable detectors.

The other kind of experiment involves search for astronomical sources of radiation. This also requires development of a detector to measure the Riemann tensor.

Measurement of the Riemann tensor. A single mass does not give information because the observer undergoes the same motion, as required by the equivalence principle. If two masses are used we may observe relative displacements.

A practical device is tremendously improved if nongravitational forces are made use of as restoring forces. The simplest detector would consist of two masses and a spring:



If the relative displacement is denoted by ξ^μ and if the (sufficiently good) approximation is made that the center of mass world line of the oscillator is a geodesic with four velocity U^α, then the equation of motion may be shown to be

$$\delta^2\xi^\mu/\delta s^2 + R^\mu{}_{\alpha\beta\gamma}\, U^\alpha r^\beta U^\gamma = f^\mu/mc^2.$$

Here f^μ is the nongravitational force acting on particle 2, and m is the reduced mass of the oscillator. In a comoving Fermi coordinate system for a harmonic oscillator this becomes

$$\frac{d^2\xi^\mu}{dt^2} + \frac{D}{m}\frac{d\xi^\mu}{dt} + \frac{k\xi^\mu}{m} = -c^2 R^\mu{}_{0\alpha 0} r^\alpha,$$

where r is the equilibrium length of the spring, k is the elastic constant, and D is the damping constant. Similar equations have been deduced for the propagation of waves in an elastic solid. For one-dimensional elastic waves, using a Riemann normal coordinate system, we have for the strain θ

$$Y\frac{\partial^2\theta}{\partial x^2} - \rho\frac{\partial^2\theta}{\partial t^2} - b\frac{\partial\theta}{\partial t} = c^2\rho R^1{}_{010}.$$

It is clear from these equations that a harmonic oscillator or the normal modes of an elastic solid may be used to measure the Riemann tensor and search for gravitational radiation.

An important question is what is the absorption cross section of a gravitational wave antenna.

We recall that the absorption cross section of an electromagnectic wave antenna averaged over all orientations in space is independent of the detailed structure of the antenna and is equal to a wavelength squared divided by 4π. This result is based in part on the fact that the radiation resistance of the antenna can be matched by an equal load resistance. If the latter condition could be met in gravitation theory a similar result would follow. Unfortunately, the gravitational radiation resistance is so incredibly small that other processes which produce dissipation of energy within the antenna give rise to resistance many orders larger than the gravitational radiation resistance. Under these conditions the absorption cross section depends very much on the parameters of the antenna and is given by

$$\sigma = 15\pi^3 G(mr^2)\,\tau/2c\lambda^2.$$

Here σ is the cross section, λ is the wavelength, τ is the relaxation time—the time required for free oscillations to decay by a factor e. The cross section is seen to be proportional to the antenna quadrupole moment. A somewhat more revealing way of writing the cross section is the following:

$$\sigma = \frac{15\pi^3}{2}\left[\frac{Gm/c^2}{\lambda}\right]\left[\frac{c\tau}{\lambda}\right] r^2.$$

This indicates that the gravitational cross section is the optical cross section multiplied by two factors. One is the ratio of the gravitational length to the wavelength, the other is the number of wavelengths contained within the averaging length $c\tau$.

University of Maryland detector. At the University of Maryland, Forward, Zipoy and I have constructed apparatus for measurement of the Riemann tensor and search for gravitational radiation in the vicinity of the angular frequency $\omega = 10^4$. This consists of a $1\frac{1}{2}$ ton aluminum cylinder, 2 feet in diameter and 5 feet long. It is suspended in a vacuum chamber and rests on acoustic filters. Piezoelectric crystals convert the normal mode vibrations into electric fields. Cryogenic electronics achieve the best possible impedance match and give good noise performance. The sensitivity is limited by the Brownian motion, which can be estimated using the equipartition theorem, $m\omega^2\langle x^2\rangle/2 = kT/2$. This implies an ultimate strain sensitivity roughly parts in 10^{16} for observations over the averaging time of 30 seconds. This sensitivity has been achieved. Possible sources which might be observed by this apparatus are collapsing double neutron stars as suggested by F. J. Dyson and R. Forward, and collapsing supernova cores as suggested by S. Colgate.

We remarked earlier that the absorption cross section for such an antenna is proportional to its quadrupole moment. What is the largest quadrupole moment which may be achieved? The earth itself is an elastic body, and some of its normal modes have quadrupole symmetry. A number of methods may be employed to observe earth strains. We have been employing a gravimeter to infer the earth strains by observing the time dependent acceleration due to gravity of the earth's surface. Strains of about a part in 10^9 may be observed. Fourier analysis has thus far failed to reveal that the modes of quadrupole symmetry of the earth are

being excited in preference to the modes of spherical symmetry. Seismic and meteorological activity limits the sensitivity of the earth as a detector. Apparatus is being developed for emplacement on the moon, by one of the Apollo Astronauts. The earth's 54 minute mode has an absorption cross section of roughly 10000 square meters, and the 15 minute lunar mode has an absorption cross section of roughly 100 square meters.

New laboratory sources. The kind of detectors we have discussed represent very substantial improvements over earlier available devices. Improvement has also been achieved in sources for dynamical gravitational fields. The spinning rod is not a good source because ultimately rupture occurs due to concentration of stresses in the vicinity of one point, the center. Improvement can be achieved by making use of the stresses in a rectangular acoustic resonator, excited in one of its vibrational modes. If rupture occurs it is now due to stresses over a plane. The velocity of sound is five orders less than the velocity of light. To obtain a radiator with linear dimensions about a gravitational wavelength, an acoustic standing wave system composed of many acoustic wavelengths may result. Ordinarily each half acoustic wavelength section oscillates out of phase with that of its nearest neighbor. The result may be an assemblage of quadrupoles with a total radiation no larger than a single small quadrupole because of the radiation cancellation from alternate half acoustic wavelengths. By use of the piezoelectric effect it is possible to build up large polarization charges on the end faces of a crystal. Under certain conditions of operation these result in stress components which do not alternate in sign every half acoustic wavelength. By such means it is possible then to achieve a source in which large stresses are maintained over volume regions with extensions of the order of gravitational wavelengths. Thus by the combination of new sources and detectors we have substantially improved the outlook for Hertz type of laboratory experiments.

UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND

Roger Penrose

Conserved Quantities and Conformal Structure in General Relativity

1. **Conserved quantities in the presence of gravitational radiation.** In the linearized theory of general relativity, gravitational radiation is accompanied by the conservation of mass, momentum and angular-momentum for regions surrounded by empty space. However, as was first shown by Bondi [1] (cf., also [2], [9], [5]), mass loss will accompany outgoing gravitational radiation in the full theory. There are, nevertheless, for asymptotically flat space-times in the presence of gravitational radiation, ten *new* conserved quantities [4]. The sketch (Figure 1) shows how these and mass might be measured.

Here m_1 is the mass of the source before radiation and m_2 is the mass of the source after radiation. We may measure m_1 by examining the field at large distances from the source, on a space-like hypersurface which cuts the world tube of the body before it emits the radiation. If we try to measure m_2 similarly on an ordinary spacelike hypersurface which cuts this world tube after the radiation is emitted, then we find that the hypersurface must intercept the *radiation* also. Thus, instead of measuring m_2 we will be measuring

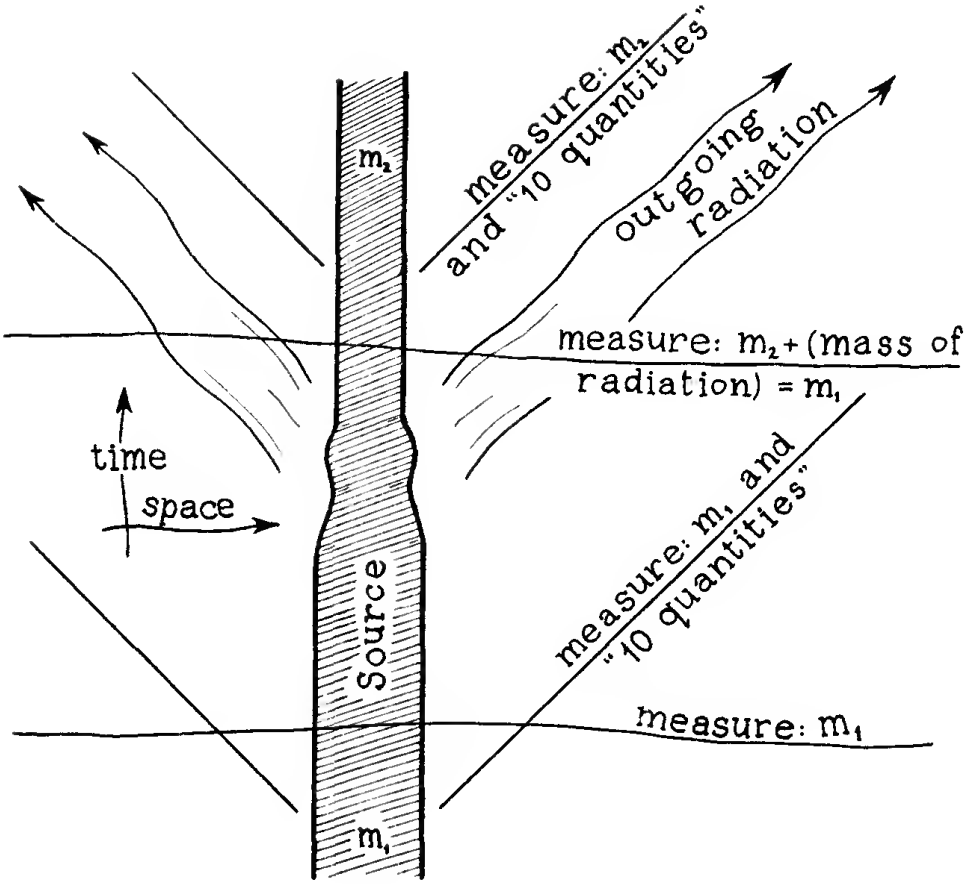


FIGURE 1

m_2 plus the mass carried by the radiation, which is again equal to m_1. We can avoid this difficulty if we use *null* hypersurfaces on which to measure mass. The mass m_2 can be measured if we choose a null hypersurface which lies entirely within the cone of radiation. We can similarly measure m_1 on a null hypersurface which opens out into the future before the radiation is emitted.

Similarly to the case of mass, the ten new quantities can be measured at infinity on any null hypersurface opening into the future, but unlike the case of mass they turn out to be the *same* whichever hypersurface is chosen provided the space-time is empty at infinity. Their exact definition will be given later.

The ten quantities can also be defined in the linear theory, but then it turns out that they always vanish for retarded fields.

For any *stationary* vacuum field (with asymptotic flatness) the ten quantities can be expressed schematically as

(1) (mass) × (quadrupole moment) − (dipole moment)^2.

(We will not prove this here, however.) The definition of these multipole moments can be taken from linear theory. In the full theory, both the source *and* its field must be expected to contribute to the values of the moments, but in any case, the quadrupole term measures an *asymmetry* of the source. The dipole moment includes an imaginary part describing the angular momentum. The expression is origin independent (as it must be since, if we use a null cone on which to measure the ten quantities, we can choose any origin as its vertex).

Let us do a simple thought experiment. We start with two hemispheres with their flat surfaces held apart by a small rod, as a stationary solution in asymptotically flat space-time. Then in a finite time interval this configuration is allowed to collapse into a sphere. As the gravitational field of the body changes, this is accompanied by a disturbance travelling outwards with the speed of light, so the body radiates gravitationally. It is easy to arrange things so that the source does not bounce (Figure 2).

Now, the ten quantities must be conserved. If our null hypersurface is chosen correctly, our measurement of the quadrupole and dipole moments of the sphere can surely be made zero so that the ten quantities would apparently vanish. But the ten quantities do *not* all vanish for our *original* configuration. Clearly this shows that the original configuration cannot be deformed into an *exact* stationary Schwarzschild solution under these conditions. Thus, it must be that more terms, which have to be added to the expression (1) to deal with *nonstationary* cases, are relevant here, so that the resulting ten quantities can be conserved. These terms perhaps involve an incoming field (since in the linear theory the ten quantities are nonzero only for fields with an incoming component). Here, this field would have to come from back-scattering of the gravitational waves, since we are supposing that no waves come in from infinity.

Note that in the Schwarzschild solution the curvature behaves as the mass times r^{-3}, so that we expect to measure mass in general by examining r^{-3} terms in the curvature. On the other hand, the ten quantities are obtained from r^{-6} terms in the curvature. The back-scattering presumably decreases locally with time even though the r^{-6} part of the back-scattered field does not decrease. One can think of the final state as being asymptotically spherical.

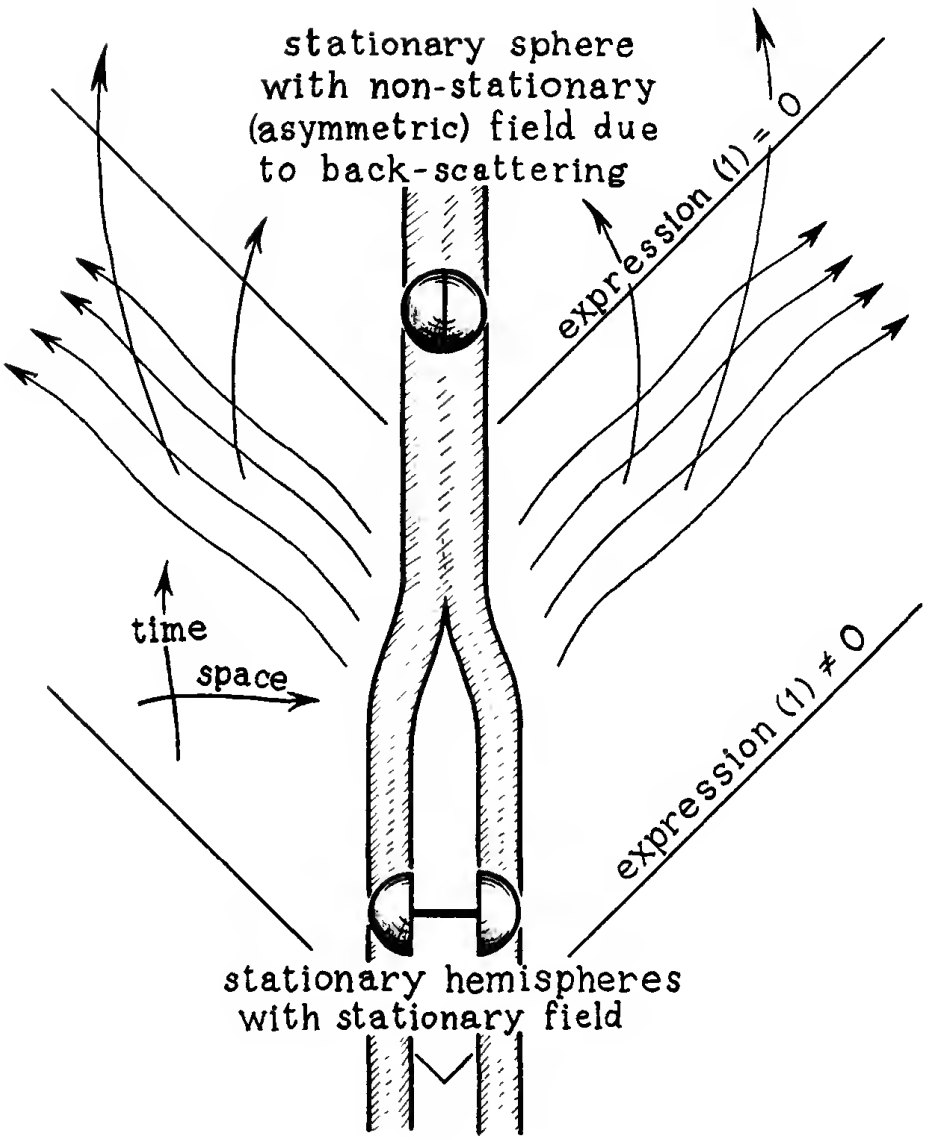


FIGURE 2. Back-scattering

2. **Conformal technique.** We wish to measure quantities at infinity and to set boundary conditions at infinity which insure the finiteness of various physical quantities (such as total energy, etc.). If we examine space-time from the point of view of its conformal structure only (e.g., angles, null cones, null geodesics), we can discuss infinity without recourse to an asymptotic limit. It turns out that we can, for asymptotically flat space-times, introduce two null hypersurfaces at infinity (past and future) which can be treated, con-

formally, the same way as finite hypersurfaces. Null infinity of the original space is represented by these hypersurfaces [6], [7].

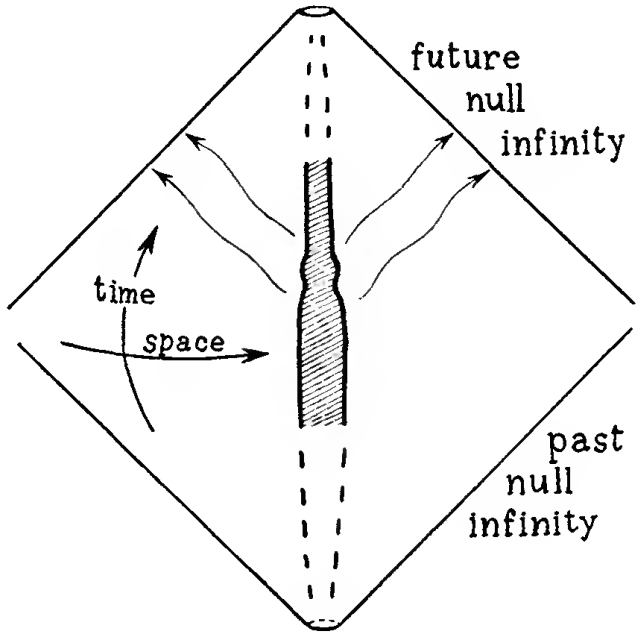


FIGURE 3

Let us consider the simplest case—Minkowski space-time $\tilde{M}_4$, where

$$d\tilde{s}^2 = du\,dv - (u - v)^2(d\theta^2 + \sin^2\theta d\phi^2)/4,$$

$$v = t + r, \qquad u = t - r.$$

We introduce a conformal transformation with conformal factor

$$\Omega = (1 + u^2)^{-1/2}(1 + v^2)^{-1/2}.$$

Here, tildes will always refer to quantities in the original physical space. The transformed space is a conformally equivalent, "non-physical" space M_4 with metric ds.

We put

$$ds = \Omega\, d\tilde{s},$$

and, since $\Omega \to 0$ at infinity, distance will become infinitely compressed at infinity. Now,

$$ds^2 = \Omega^2 d\tilde{s}^2 = dp\,dq - \sin^2(p - q)(d\theta^2 + \sin^2\theta d\phi^2)/4$$

with $u = \tan p$ and $v = \tan q$ (so that points at infinity are not given infinite coordinates). The two hypersurfaces representing infinity are given by $q = \pi/2$, $-\pi/2 < p < \pi/2$ and $p = -\pi/2$, $-\pi/2 < q < \pi/2$. These hypersurfaces represent the future and past null infinities. In addition there are three points I^- $(p = -\pi/2,\ q = -\pi/2)$, I^0 $(p = -\pi/2,\ q = \pi/2)$, and I^+ $(p = \pi/2,\ q = \pi/2)$ representing past timelike, spacelike and future timelike infinities. (I^0 is, in fact, a point, but this is not important here.) Space is regular in the neighborhood of the bounding hypersurface in M_4.

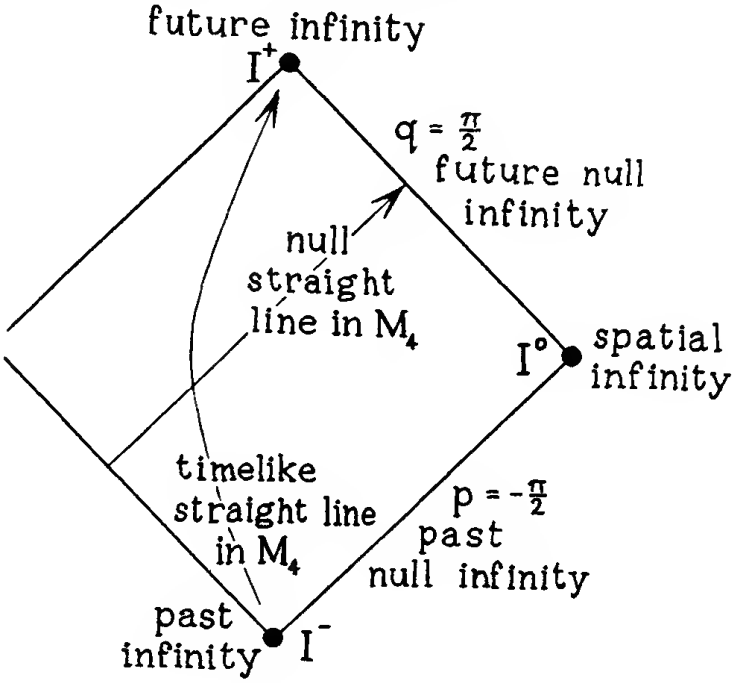


FIGURE 4. Finite M_4

Our definition of asymptotic flatness in curved space-time $\tilde{M}$ with $G_{\mu\nu} = 0$ at infinity is that M exist, with interior conformally equivalent to $\tilde{M}$ via $ds = \Omega d\tilde{s}$, so that:

(b) Ω suitably differentiable throughout M with $\Omega = 0$ on the boundary $\dot{M}$ of M,

(c) suitable topological conditions (e.g., null geodesics have two end points on $\dot{M}$).

The metric

$$d\tilde{s}^2 = R^{-2}AdR^2 + 2B_i dx^i dR + R^2 C_{ij} dx^i dx^j,$$

where $i, j = 1, 2, 3$ and A, B_i and C_{ij} are sufficiently differentiable

functions of $x^\mu (\mu = 0, 1, 2, 3)$, with $x^0 = R^{-1}$, includes as special cases many metrics which have been suggested as representing gravitational radiation [2], [8], [9] (and probably also [5]). If

$$\begin{vmatrix} A & B_j \\ B_i & C_{ij} \end{vmatrix} \neq 0,$$

this implies that

$$ds = \Omega d\tilde{s},$$

with $\Omega = R^{-1}$, is regular on our bounding hypersurface of M. (These metrics also include de Sitter space and the asymptotically de Sitter spaces of Demianski. These are not asymptotically flat in the usual sense, even though (a), (b), (c) are satisfied. This is because, here, $G_{\mu\nu} = -\lambda g_{\mu\nu} \neq 0$.)

3. **The field at infinity and the conserved quantities.** A zero rest-mass free-field can be regarded as being conformally invariant. For a spin two field consider a tensor $K_{\mu\nu\rho\sigma}$ subject to

$$\begin{aligned} K_{\alpha\beta\rho\sigma} &= K_{[\rho\sigma][\alpha\beta]}, \\ K_{\alpha[\nu\rho\sigma]} &= 0, \\ K^{\alpha}{}_{\rho\alpha\sigma} &= 0 \end{aligned} \tag{2}$$

and the "free-field equation"

$$\nabla_{[\mu} K_{\alpha\beta]\rho\sigma} = 0, \tag{3}$$

where ∇_μ indicates covariant differentiation (and the square brackets denote skew-symmetrization). In flat space, (2) and (3) is the gauge invariant description of a spin two, zero rest-mass free-field (i.e., linearized source-free gravitation). In curved space, the Equations (2) and (3) can still be used, but this implies a relation[1] connecting $K_{\mu\nu\rho\sigma}$ with the conformal curvature that must be satisfied. Equation (3) is preserved under the conformal transformation

$$g_{\mu\nu} = \Omega^2 \tilde{g}_{\mu\nu} \tag{4}$$

if

$$K_{\mu\nu\rho\sigma} = \Omega \tilde{K}_{\mu\nu\rho\sigma}. \tag{5}$$

[1] The general relation is $K_{\mu\nu\rho[\sigma} C_{\tau]}{}^{\mu\nu\rho} = 0$.

The conformal tensor $C_{\mu\nu\rho\sigma}$ satisfies (2), and also (3), whenever $G_{\mu\nu} = 0$. (Actually weaker conditions will suffice.) Thus, in the *original* space we can set

$$\widetilde{K}_{\mu\nu\rho\sigma} = \widetilde{C}_{\mu\nu\rho\sigma}, \tag{6}$$

since this will imply that (3) holds in $\widetilde{M}$, in the regions free of matter. However,

$$C_{\mu\nu\rho\sigma} = \Omega^2 \widetilde{C}_{\mu\nu\rho\sigma}$$

(cf., Schild, *Lectures on general relativity theory,* these Lectures). Thus, in the unphysical space this entails

$$C_{\mu\nu\rho\sigma} = \Omega K_{\mu\nu\rho\sigma}. \tag{7}$$

We now have the free-field Equations (3) holding for $K_{\mu\nu\rho\sigma}$ (and not $C_{\mu\nu\rho\sigma}$) in the unphysical space M, in the regions corresponding to the empty regions of $\widetilde{M}$. We wish to study the asymptotic gravitational field[2] by examining $K_{\mu\nu\rho\sigma}$ on the boundary hypersurface $\dot{M}$. For this to work, $K_{\mu\nu\rho\sigma}$ must be regular and finite on $\dot{M}$. But from (7), this clearly could only happen if $C_{\mu\nu\rho\sigma} = 0$ on $\dot{M}$. Fortunately, it can be shown that our assumptions already imply $C_{\mu\nu\rho\sigma} = 0$ there, and it follows that $K_{\mu\nu\rho\sigma}$ is *necessarily* regular on $\dot{M}$.

We will measure the energy-momentum and the ten conserved quantities on a null hypersurface $\widetilde{S}$, opening into the future, since we are considering outgoing radiation. Let S be the image of $\widetilde{S}$ in M. We are concerned with measurements made at infinity, so it is really only the quantities defined on the intersection Σ of $\dot{M}$ with S that will concern us. Suppose, now, we wish to compare a quantity measured on one hypersurface $\widetilde{S}_1$ with the same quantity measured on a "later" hypersurface $\widetilde{S}_2$. Let Σ_1 and Σ_2 be the respective intersections of M with S_1 and S_2 (Figure 5). We will have an integral over Σ_1 which we wish to compare with a corresponding integral over Σ_2. The idea here is to use Gauss' theorem on the three-surface $\dot{M}$. The difference between the integrals over Σ_1 and Σ_2 will then be expressed as an integral of some "divergence" over

[2] The point of view here is that $K_{\mu\nu\rho\sigma}$, subject to (2) and (3), is the zero rest-mass spin two field which describes "free gravitation," this "free gravitational field" being conformally invariant. The "interactions" are expressed through the Bianchi and Ricci identities as applied to $C_{\mu\nu\rho\sigma} = \Omega K_{\mu\nu\rho\sigma}$. The conformal *noninvariance* of these "interactions" is manifested in the explicit appearance of Ω in these expressions.

the portion M' of M lying between Σ_1 and Σ_2. If this divergence vanishes, we will have an exactly conserved quantity. Otherwise, we may be able to regard the integral over $\dot{M}'$ as the contribution to the quantity in question which is carried away by the gravitational radiation.

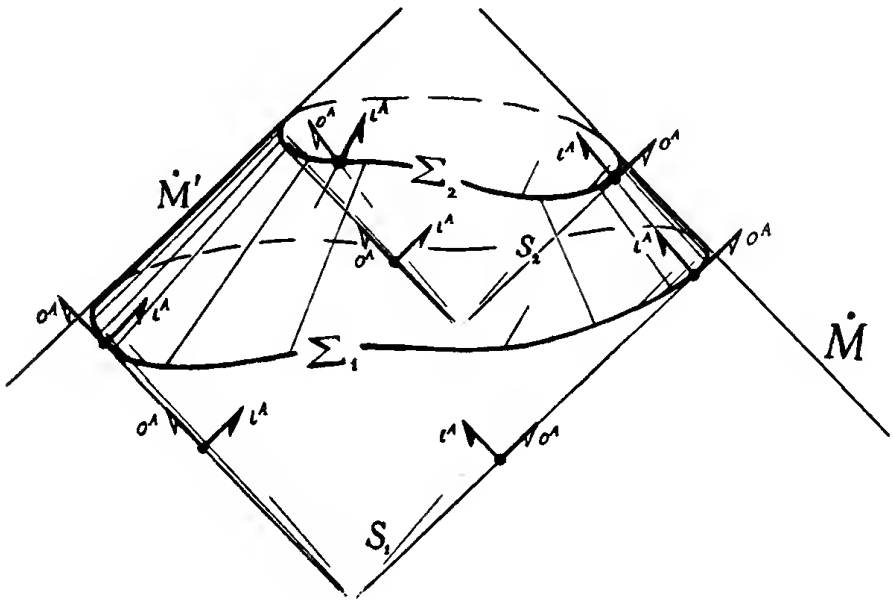


FIGURE 5

For asymptotically flat spaces, $\dot{M}$ is in fact a null hypersurface [7]. The structure of $\dot{M}$ is essentially the same as for Minkowski space (Figure 4). We shall omit the three points I^-, I^0, I^+ here. Then $\dot{M}$ consists of two portions, each of which is topologically a "cylinder" $S^2 \times \mathrm{E}^1$. We are concerned, here, only with the future portion $\dot{M}^+$, and by judicious choice of conformal factor Ω, we can ensure that the geometry of $\dot{M}^+$ is as simple as possible. In fact, by taking one generator of $\dot{M}^+$ "back to infinity" we can open out the cylinder into a space with Euclidean three-space topology. Furthermore, it turns out that we can also make this three-space metrically *flat* (Figure 6). This will simplify matters considerably.

In order to proceed further, we will need to use two-component spinors (cf., Schild, *Lectures on general relativity theory,* these Lectures). The quantity $\sigma_\mu{}^{AB'}$ (Hermitian (2×2) matrix for each μ) translates tensor quantities into spinor quantities and satisfies

$$\sigma_\mu{}^{AB'} \sigma_\nu{}^{CD'} \epsilon_{AC} \epsilon_{B'D'} = g_{\mu\nu}.$$

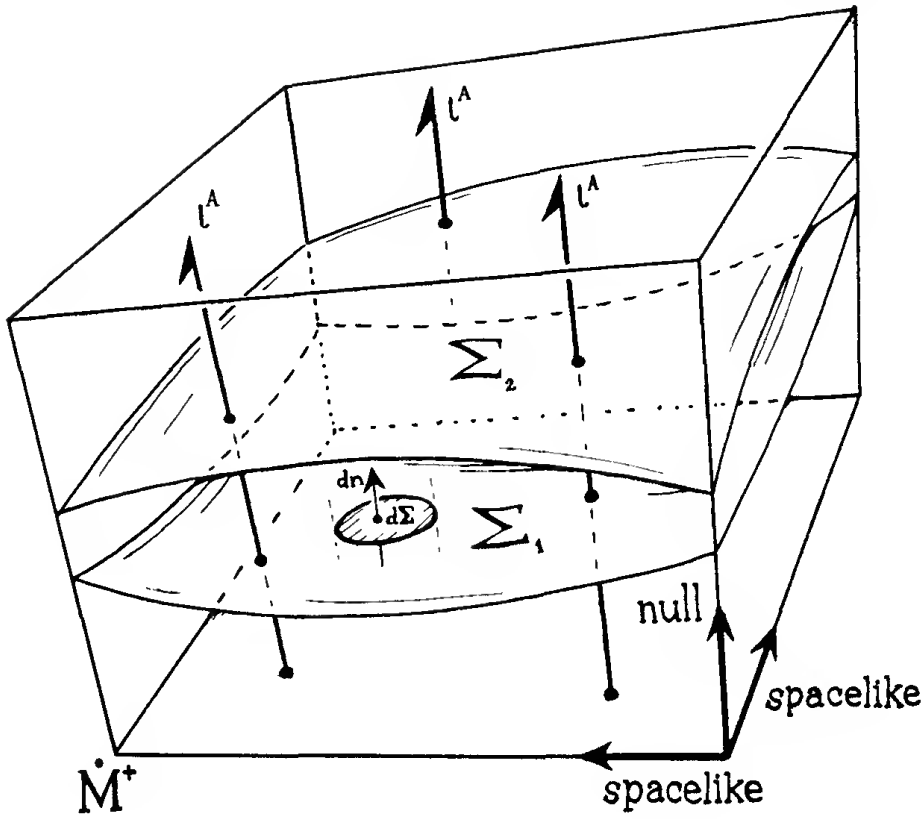


FIGURE 6

To preserve this under conformal transformation we set

$$\sigma_\mu{}^{AB'} = \Omega\tilde{\sigma}_\mu{}^{AB'} \tag{8}$$

with

$$\epsilon_{AB} = \tilde{\epsilon}_{AB}, \quad \epsilon_{A'B'} = \tilde{\epsilon}_{A'B'}.$$

We have

$$K_{\mu\nu\rho\sigma} \leftrightarrow K_{AP'BQ'CR'DS'} = \phi_{ABCD}\epsilon_{P'Q'}\epsilon_{R'S'} + \epsilon_{AB}\epsilon_{CD}\bar{\phi}_{P'Q'R'S'},$$

$$C_{\mu\nu\rho\sigma} \leftrightarrow C_{AP'BQ'CR'DS'} = \Psi_{ABCD}\epsilon_{P'Q'}\epsilon_{R'S'} + \epsilon_{AB}\epsilon_{CD}\bar{\Psi}_{P'Q'R'S'},$$

so that

$$\phi_{ABCD} = \Omega^{-1}\Psi_{ABCD} = \Omega^{-3}\tilde{\phi}_{ABCD} = \Omega^{-3}\tilde{\Psi}_{ABCD} \tag{9}$$

by (4), (5), (6), (7), (8).

We introduce a spinor reference dyad o^A, ι^A in the neighborhood of $\dot{M}^+$, with $o_A\iota^A = 1$ [3], where the null direction in $\dot{M}^+$ is represented by ι^A and the null directions in S_1 and S_2 are represented by o^A (Figure 5). It turns out we can choose ι^A and Ω so that

$$\left.\begin{array}{ll} \nabla_{AB'}\iota_C = 0, & o^A\bar{o}^{B'}\nabla_{AB'}o_C = 0 \\ \Psi_{ABCD} = 0, & R = 0 \\ R_{AC'BD'}\iota^A\iota^B = -2N\bar{\iota}_{C'}\bar{\iota}_{B'}, & \end{array}\right\} \text{on } \dot{M}^+ \tag{10}$$

where R is the scalar curvature and where $R_{AC'BD'}$ and $\nabla_{AB'}$ are, respectively, the spinor translation of the Ricci tensor $R_{\mu\nu}$ (which generally does not vanish in M) and of the covariant derivative operator ∇_μ. These equations will be required shortly.

Let $\tilde{\iota}^A$, $\tilde{o}^A$ and r be defined on $\tilde{M}$ so that the null directions in S_1 and S_2 are represented by $\tilde{o}^A$. Let both $\tilde{\iota}^A$ and $\tilde{o}^A$ be parallelly propagated along the generators of $\tilde{S}_1$ and $\tilde{S}_2$, and so that (near $\dot{M}^+$) ι^A, o^A are the respective images of $\tilde{\iota}^A$, $\tilde{o}^A$ in M. (This can always be arranged.) The radial parameter r is an affine parameter on each generator of $\tilde{S}_1$ and of $\tilde{S}_2$ (scaled so that $\tilde{o}^A\bar{\tilde{o}}^{B'}\nabla_{AB'}r = 1$). We have [7]

$$\tilde{o}^A \sim \Omega^{1/2}o^A, \qquad \tilde{\iota}^A \sim \Omega^{-1/2}\iota^A, \qquad r \sim \Omega^{-1}. \tag{11}$$

The five complex components of $\tilde{C}_{\mu\nu\rho\sigma}$ may be defined by

$$\begin{aligned} \Psi_0 &= \tilde{\Psi}_{ABCD}\tilde{o}^A\tilde{o}^B\tilde{o}^C\tilde{o}^D = O(r^{-5}), \\ \Psi_1 &= \tilde{\Psi}_{ABCD}\tilde{o}^A\tilde{o}^B\tilde{o}^C\tilde{\iota}^D = O(r^{-4}), \\ \Psi_2 &= \tilde{\Psi}_{ABCD}\tilde{o}^A\tilde{o}^B\tilde{\iota}^C\tilde{\iota}^D = O(r^{-3}), \\ \Psi_3 &= \tilde{\Psi}_{ABCD}\tilde{o}^A\tilde{\iota}^B\tilde{\iota}^C\tilde{\iota}^D = O(r^{-2}), \\ \Psi_4 &= \tilde{\Psi}_{ABCD}\tilde{\iota}^A\tilde{\iota}^B\tilde{\iota}^C\tilde{\iota}^D = O(r^{-1}). \end{aligned}$$

The order of magnitude behavior of these quantities is a consequence of the asymptotic flatness of the space ("peeling-off" property [3], [9]) and, more specifically, of the finiteness of $K_{\mu\nu\rho\sigma}$ (i.e., of ϕ_{ABCD}) on $\dot{M}^+$. This follows from (9) and (11) [7]. Put

$$\Psi_0 = \Psi_0^0 r^{-5} + \Psi_0^1 r^{-6} + O(r^{-7}), \tag{12}$$

$$\Psi_2 = \Psi_2^0 r^{-3} + O(r^{-4}). \tag{13}$$

By writing $\alpha_A = \iota^B\alpha_{AB}$, where α_{AB} is a symmetric spinor representing a bivector, we can put Gauss' theorem (applied to this bivector) into the form

$$\int_{\Sigma_1}\alpha_A o^A d\Sigma_2 - \int_{\Sigma_2}\alpha_A o^A d\Sigma_1 = \int_{\dot{M}'}\bar{\iota}^{C'}\nabla^A_{C'}\alpha_A\, d\Sigma\, dn,$$

where $d\Sigma_2$, $d\Sigma_1$, $d\Sigma$ are surface area elements ($d\Sigma$ referring to the

area of a cross-section of M'—the portion of M^+ between Σ_1 and Σ_2) and dn is an element of affine distance in the null direction in M^+ (Figure 6). For the ten conserved quantities we take

$$\bar{\alpha}_{P'} = o^A o^B o^C o^D o^E \nabla_{EP'} \phi_{ABCD} + 5 o^A o^B o^C o^D \phi_{ABCD} \nabla_{QP'} o^Q.$$

Then it turns out (from (10) and (3)) that

$$\bar{\iota}^{B'} \nabla_{AB'} \alpha^A = 0,$$

so that

$$\int_{\Sigma_1} \alpha_A o^A d\Sigma_1 = \int_{\Sigma_2} \alpha_A o^A d\Sigma_2.$$

Here

$$\bar{\alpha}_{P'} \bar{o}^{P'} = o^A o^B o^C o^D (o^E \bar{o}^{P'} \nabla_{EP'} + 5 \bar{o}^{P'} \nabla_{QP'} o^Q) \phi_{ABCD}. \tag{14}$$

In the physical space $\widetilde{M}$ this amounts to the term Ψ_0^1 of (12). (The $o^E \bar{o}^{P'} \nabla_{EP'}$ derivative acts along the generators of $\widetilde{S}_1$, $\widetilde{S}_2$ and takes us to the r^{-6} term. The term with $5\bar{o}^{P'} \nabla_{QP'} o^Q$ is a correction term to make the operation conformally invariant, and it knocks out the r^{-5} term of (12).) Thus, roughly speaking, if we integrate the Ψ_0^1 for $\widetilde{S}_1$, we get the same answer as for $\widetilde{S}_2$; however, we must take into account the weighting factors involved in the conformal transformation. In fact, we get different weighting factors depending upon which generator of $\dot{M}^+$ has been chosen to be "at infinity." Because of the conformal properties of (14), it turns out that there are just five linearly independent such weighting factors. The final result can be stated [4]:

$$Q_m = \int \Psi_0^1 {}_2\overline{Y}_{2,m} \sin\theta \, d\phi \, d\theta \qquad (m = -2, -1, 0, 1, 2),$$

where $r^2(d\theta^2 + \sin^2\theta \, d\phi^2)$ defines the metric of the sphere at infinity on $\widetilde{S}_1$ or $\widetilde{S}_2$. The "weighting factors" ${}_2Y_{2,m}$ are "tensor" spherical harmonics defined by

$${}_2Y_{2,m} = 24^{-1/2} \left(\sin\theta \frac{\partial}{\partial\theta} + i \frac{\partial}{\partial\phi} \right) \left(\frac{1}{\sin\theta} \frac{\partial}{\partial\theta} + \frac{i}{\sin^2\theta} \frac{\partial}{\partial\phi} \right) Y_{2,m}$$

where $Y_{2,m}$ are ordinary harmonics. The real and imaginary parts of the five Q_m are then the required conserved quantities.

In the same way, we can treat energy-momentum. Here, put

$$\alpha_A = -\phi_{ABCD}o^B\bar{\iota}^C\iota^D + N\bar{\iota}^{C'}o^B\nabla_{AC'}o_B$$

(for N, see (10)). It then turns out that

$$\bar{\iota}^{C'}\nabla^A_{C'}\alpha_A = |N|^2.$$

The quantity N is essentially the Bondi-Sachs "news function," with $|N|^2$ representing the energy-momentum flux of the gravitational waves at infinity. We have

$$\alpha_A o^A = -\phi_{ABCD}o^A o^B \bar{\iota}^C \iota^D + N\bar{\iota}^{C'}o^A o^B \nabla_{AC'}o_B$$

which, when translated into the physical space $\tilde{M}$, becomes $-\Psi_2^0 + N\sigma^0$ (cf., (13)). The term $N\sigma^0$ is essentially a correction term arising from the fact that the hypersurfaces $\tilde{S}_1$, $\tilde{S}_2$ need not be as "isotropic" asymptotically as ordinary null cones in Minkowski space. The term σ^0 measures the asymptotic shear of $\tilde{S}_1$, $\tilde{S}_2$. Again the conformal transformation leads to different choices of weighting factors arising. Here there are four linearly independent choices—corresponding to "total energy" and the three components of "total linear momentum." The integral of $-\Psi_2^0$, "corrected" by $N\sigma^0$, over the sphere at infinity of $\tilde{S}_1$ or $\tilde{S}_2$ then gives the total mass of the system intercepted by the hypersurface.

References

1. H. Bondi, Nature 186 (1960), 535.
2. H. Bondi, M. G. J. van der Burg and A. W. K. Metzner, Proc. Roy. Soc. Ser. A 269 (1962), 21.
3. E. T. Newman and R. Penrose, J. Math. Phys. 3 (1962), 566; 4 (1963), 998.
4. ———, Phys. Rev. Lett. 15 (1965), 231.
5. E. T. Newman and T. W. J. Unti, J. Math. Phys. 3 (1962), 891.
6. R. Penrose, Lectures, Les Houches, 1963 Summer School of Theoret. Phys., Univ. Grenoble, pp. 563-584. Gordon and Breach, New York, 1964.
7. ———, Proc. Roy. Soc. Ser. A 284 (1965), 159.
8. I. Robinson and A. Trautman, Proc. Roy. Soc. Ser. A 265 (1962), 463.
9. R. Sachs, Proc. Roy. Soc. Ser. A 270 (1962), 103.

BIRKBECK COLLEGE, UNIVERSITY OF LONDON
LONDON, ENGLAND

C. W. Misner

Taub-NUT Space as a Counterexample to Almost Anything

This lecture will discuss some of the peculiar properties of the metric

$$ds^2 = (t^2 + l^2)(d\theta^2 + \sin^2\theta d\phi^2) + U(t)(2l)^2(d\psi + \cos\theta d\phi)^2 + 2(2l)(d\psi + \cos\theta d\phi)\, dt \tag{1}$$

where

$$U(t) = -1 + 2(mt + l^2)/(t^2 + l^2). \tag{2}$$

This metric satisfies the empty-space Einstein equations

$$R_{\mu\nu} = 0 \tag{3}$$

and has been discovered by both of the prime exact-solution-finding methods mentioned by Sachs in his paper in these Lectures. Taub [5] discovered it in a systematic development of a class of metrics with high symmetry. Later it was rediscovered by Newman, Tamburino and Unti [4] studying a class of algebraically special metrics. Actually, Taub gave the metric in a coordinate system covering only the region where $U(t) > 0$ ("Taub space") in which the $t = \text{const}$ hypersurfaces are spacelike, while Newman, Unti and Tamburino gave the region where

$U(t) < 0$ ("NUT space") in which the ψ-lines ($t\theta\phi$ constant) are time-like.

This Taub-NUT space has many unusual properties, some of which are also known in other metrics. I will give a short list and then discuss a few:

1. Although the Taub region $U > 0$ can be interpreted as a cosmological solution with homogeneous but nonisotropic space sections, it evolves into NUT space which seems to have no reasonable interpretation.

2. The NUT region contains closed time-like lines.

3. The NUT region does not contain any decent space-like hypersurfaces.

4. Although the curvature tensor vanishes as one approaches infinity in space-like directions, asymptotically rectangular coordinates ($g_{\mu\nu} \to \eta_{\mu\nu}$) do not exist.

5. Taub space allows, besides Equation (1), another, inequivalent, maximal analytic extension.

6. Taub-NUT space is nonsingular in a meaningful mathematical sense but is not geodesically complete.

7. There are closed geodesics (circles) on which one cannot extend the solution of the geodesic equation to infinite values of the affine path parameter.

The first four points are discussed in my paper [3] on NUT space and will not be considered further here. Points 6 and 7 are based on a study of the geodesics which Taub and I [6] hope to finish writing soon. An example of inextendable closed geodesics (7) for a simpler metric is given in the paragraph containing Equation (1) in my paper [3]. The question of the nonuniqueness of analytic continuation for metrics (point 5) is also based on the paper by Taub and Misner [6] and will occupy us for the remainder of this lecture.

Before discussing analytic continuations we must discuss analyticity. A function of several real variables $f(xy \cdots z)$ is analytic at $x_0 y_0 \cdots z_0$ if it has a power series expansion about that point with a nonzero radius of convergence. A function $f(P)$ on a manifold M is analytic if it can be represented as an analytic function $f(x^1, x^2, \cdots, x^4)$ of the coordinates in the various coordinate patches defining the analytic structure of the manifold (see Auslander and MacKenzie [1] for the definition of manifolds and of differentiable

structure). Tensors are analytic if their components are analytic functions of the coordinates in each coordinate patch. A symmetric tensor $g_{\mu\nu}$ is a metric only if it has the proper signature $(-+++)$; in particular $g = \det g_{\mu\nu}$ cannot vanish. The metric of Equation (1) fails to satisfy the signature requirement at $\theta = 0, \pi$ since in these $t\psi\theta\phi$ coordinates one has

$$(-g)^{1/2} = 2l(t^2 + l^2)\sin\theta. \tag{4}$$

We will therefore try to interpret Equation (1) as defining the metric on a coordinate patch where $t\psi\theta\phi$ have the ranges

$$\begin{aligned} &-\infty < t < +\infty, && 0 < \theta < \pi, \\ &0 < \psi < 4\pi, && 0 < \phi < 2\pi. \end{aligned} \tag{5}$$

Analyticity of the metric in this range is obvious, as is the choice of limits on t and θ. It remains to be seen whether there exists any larger manifold containing this coordinate patch on which the

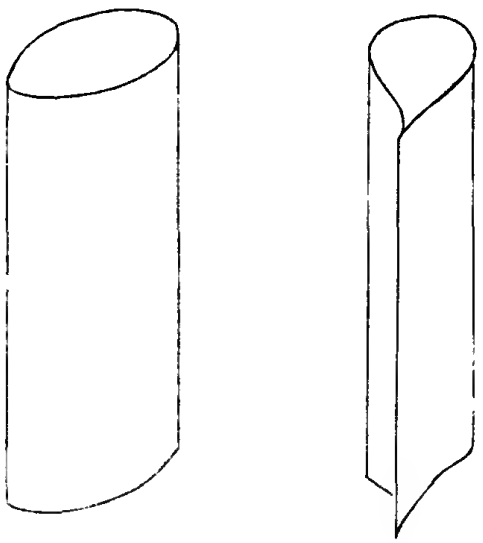


FIGURE 1. Both the cylinders shown here are flat and can be represented away from the seam ($\theta = 0$ or 2π) by the metric of Equation (6). Only (a) is a smooth (analytic) manifold, while (b) with its sharp-edged seam does not inherit a useful class of differentiable functions from the Euclidean 3-space in which it is embedded.

metric can be analytic. We will study some simpler examples before returning to this question.

As a very simple example consider the metric

$$ds^2 = dt^2 + d\theta^2 \tag{6}$$

on a coordinate patch $-\infty < t < \infty$, $0 < \theta < 2\pi$ (Figure 1). We are accustomed to interpret this metric as a cylinder $R \times S^1$. Here R means the real line $-\infty < t < \infty$, and S^1 is the 1-sphere or circle (base of the cylinder). I shall not attempt to explain how one looks at a metric on a single coordinate patch, such as Equation (1) or Equation (6), and tries to guess what the full manifold containing this patch should be. But we shall now consider how to verify a guess. I guess that the metric

$$ds^2 = (dx^2 + dy^2)/(x^2 + y^2) \tag{7}$$

which is clearly analytic in the region $0 < x^2 + y^2 < +\infty$ might be related to the metric of Equation (6). The coordinate transformation

$$x = e^t \cos\theta, \qquad y = e^t \sin\theta \tag{8}$$

shows indeed that these metrics are equivalent on the region $\theta \neq 0$ (which is the largest region to which this correspondence between xy and $t\theta$ coordinates can be extended in a one-to-one way) but Equation (7) is to be preferred since it is analytic also on the positive x-axis of the xy plane. One may begin to worry about the "singularity" of $x^2 + y^2 \to 0$ in Equation (7); since it corresponds to $t \to -\infty$ in Equation (6) it clearly is *not* a singularity. A coordinate independent way of stating this is to note that every geodesic which approaches the boundary of the manifold ($x^2 + y^2 \to 0$ or ∞) has infinite length.

Some slightly more complicated metrics to serve as further examples are

$$ds_{\mathrm{I}}^2 = dt^2 + d\theta^2 + \sin^2\theta d\phi^2, \tag{9}$$

$$ds_{\mathrm{II}}^2 = (1/3)(2 + \cos\theta)\, dt^2 + d\theta^2 + \sin^2\vartheta d\phi^2, \tag{10}$$

$$ds_{\mathrm{III}}^2 = dt^2 + e^{2t} d\theta^2 + \sin^2\theta d\phi^2. \tag{11}$$

The first metric here obviously represents a higher dimensional cylinder $R \times S^2$ which is just a linear (R) stack of ordinary spheres

(S^2), while the other two metrics represent some deformations of this first hypercylinder. What is not so obvious to the unpractised eye is that while the first two metrics are analytic, the third is not even continuous over the whole cylinder $R \times S^2$. Analyticity of the first metric (by which we now mean the existence of an analytic extension of it to cover all of $R \times S^2$) is demonstrated by writing

$$ds_{\mathrm{I}}^2 = (1/r^2)(dx^2 + dy^2 + dz^2) \tag{12}$$

with

$$r^2 \equiv x^2 + y^2 + z^2 \tag{13}$$

which is obviously analytic for $0 < r^2 < +\infty$, and which corresponds to Equation (9) by the transformation

$$\begin{aligned} x &= e^t \sin\theta \cos\phi, \\ y &= e^t \sin\theta \sin\phi, \\ z &= e^t \cos\theta. \end{aligned} \tag{14}$$

Use the same transformation on ds_{II}^2 and write

$$ds_{\mathrm{II}}^2 = ds_{\mathrm{I}}^2 + (1/3)(\cos\theta - 1)\, dt^2. \tag{15}$$

Then since $\cos\theta = z/r$ is analytic for $r^2 > 0$ and

$$dt = (1/r^2)(x\,dy + y\,dy + z\,dz)$$

is an analytic differential form in this region, the xyz components of ds_{II}^2 will also be analytic functions. The condition of proper signature is verified from

$$\begin{aligned} g_{\mathrm{II}} &= (1/3)(2 + \cos\theta) \sin^2\theta [\partial(t\theta\phi)/\partial(xyz)]^2 \\ &= (2r + z)/3r^7 > 0. \end{aligned} \tag{16}$$

But studying ds_{III}^2 in xyz coordinates leads to no interesting extension, for we have

$$ds_{\mathrm{III}}^2 = ds_{\mathrm{I}}^2 + (e^{2t} - 1)\, d\theta^2, \tag{17}$$

and from the transformation (14) one sees that

$$d\theta = \frac{z}{r^2} \frac{x\,dx + y\,dy}{(x^2 + y^2)^{1/2}} - \frac{(x^2 + y^2)^{1/2}}{r^2} dz \tag{18}$$

is *not* an analytic differential form on the z-axis where $x/(x^2 + y^2)^{1/2}$ for instance is not a continuous function. We could then ask

whether some coordinate transformation different from Equation (14) might not lead to a form of ds_{III}^2 which allowed some analytic extension beyond the region to $-\infty < t < +\infty$, $0 < \theta < \pi$, $0 < \phi < 2\pi$ where Equation (11) demonstrates analyticity. But a computation of the curvature tensor (see Appendix A of my 1963 paper for rapid computational techniques which here gave $R_{\mu\nu\alpha\beta}$ starting from Equation (11) in twelve minutes work) shows that

$$R_{ij}R^{ij} - R^2/2 = 2(e^t \sin\theta)^{-2}. \tag{19}$$

The lines of infinite curvature at $\theta = 0$ and $\theta = \pi$ are therefore natural boundaries to this space; since the metric puts these lines arbitrarily close to regular points of the space we say that this space is intrinsically singular.

An analytic extension of the line element of Equation (1) is obtained by the coordinate transformation

$$\begin{aligned} w &= e^{t/2}\cos\tfrac{1}{2}\theta\cos\tfrac{1}{2}(\phi+\psi), \\ x &= e^{t/2}\sin\tfrac{1}{2}\theta\cos\tfrac{1}{2}(\phi-\psi), \\ y &= e^{t/2}\sin\tfrac{1}{2}\theta\sin\tfrac{1}{2}(\phi-\psi), \\ z &= e^{t/2}\cos\tfrac{1}{2}\theta\sin\tfrac{1}{2}(\phi+\psi). \end{aligned} \tag{20a}$$

This transformation can also be written

$$q = e^{t/2}e^{k\phi/2}e^{i\theta/2}e^{k\psi/2} \tag{20b}$$

where q is the quaternion

$$q = w + ix + jy + kz$$

and ijk are the imaginary quaternion units which obey $k^2 = -1$, $ij = -ji = k$, etc. This transformation shows that $\psi\theta\phi$ are the "Euler angle" coordinates on S^3 which are familiar in discussions of the rotation group $SO(3)$. (See, for instance, Corben and Stehle [2], Appendix IV.)

In order to write the Taub-NUT metric in $wxyz$ coordinates it is convenient first to define several differential forms

$$\begin{aligned} \sigma_x &= 2|q|^{-2}(x\,dw - w\,dx - z\,dy + y\,dz), \\ \sigma_y &= 2|q|^{-2}(y\,dw + z\,dx - w\,dy - x\,dz), \\ \sigma_z &= 2|q|^{-2}(z\,dw - y\,dx + x\,dy - w\,dz), \\ dt &= 2|q|^{-2}(w\,dw + x\,dx + y\,dy + z\,dz), \end{aligned} \tag{21}$$

where

$$|q|^2 \equiv w^2 + x^2 + y^2 + z^2 = e^t. \tag{22}$$

These differential forms are obviously analytic on the region $0 < |q|^2 < \infty$, and from the following equation we see that the components of the transformed metric are also:

$$ds^2 = (t^2 + l^2)(\sigma_x{}^2 + \sigma_y{}^2) + U(t)(2l)^2\sigma_z{}^2 + 2(2l)\sigma_z dt. \tag{23}$$

This equation can also be written

$$ds^2 = g_{\mu\nu}\omega^\mu\omega^\nu \tag{24}$$

where

$$\begin{aligned} \omega^0 &= dt, \\ \omega^1 &= \sigma_x(t^2 + l^2)^{1/2}, \\ \omega^2 &= \sigma_y(t^2 + l^2)^{1/2}, \\ \omega^3 &= \sigma_z(2l) \end{aligned} \tag{25}$$

and

$$g_{\mu\nu} = \begin{pmatrix} 0 & 1 & & \\ 1 & U & & 0 \\ & & 1 & 0 \\ & 0 & 0 & 1 \end{pmatrix}. \tag{26}$$

In this form it is easy to verify that the signature requirement on the metric is satisfied, for we need only verify the signature of the $g_{\mu\nu}$ matrix in Equation (26) and check that the basis differential forms of Equation (25) are linearly independent. This linear independence is clear when one notes that the coefficients in Equations (21) form nonzero orthogonal vectors in the standard flat Euclidean metric on *wxyz* space. The signature of $g_{\mu\nu}$ in Equation (26) is obvious when $U(t) = 0$, and cannot change since

$$g = \det g_{\mu\nu} = -1 \neq 0. \tag{27}$$

It can be shown that the metric of Equation (23) on the domain $0 < |q|^2 < +\infty$ of *wxyz* space is maximal in the sense that it cannot be identified with a coordinate patch on an even larger connected manifold. Taub and I [6] have shown this by studying its geodesics and verifying that every geodesic arc which approaches the bound-

aries ($|q|^2 \to 0$ or $|q|^2 \to \infty$) is infinitely long as measured by any affine path parameter for the geodesic equation. In contrast, geodesics which start from points outside, but on the boundary of, a coordinate patch can enter it with finite, even arbitrarily small, changes in the path parameter.

Let us now turn to another question of analytic continuation, its uniqueness. We may and do now choose to consider Equation (23) on the region $0 < |q|^2 < \infty$ as giving one maximal analytic extension of Taub space, which is the region

$$e^{t_1} < |q|^2 < e^{t_2} \tag{28}$$

where t_1 and t_2 are the two zeros of $U(t)$. Since analytic continuation of functions on the real line is a unique process, we might expect this also to be true for analytic metric manifolds. To exhibit non-uniqueness we first introduce a new coordinate system on Taub space by the transformation

$$\psi = \Psi - \int_m^t (lU)^{-1} dt \tag{29}$$

in which we retain the old $t\theta\phi$ coordinates. The metric then becomes

$$\begin{aligned} ds^2 = {} & (t^2 + l^2)(d\theta^2 + \sin^2\theta d\phi^2) \\ & + U(t)(2l)^2(d\Psi + \cos\theta d\phi)^2 \\ & - 2(2l)(d\Psi + \cos\theta d\phi)\, dt \end{aligned} \tag{30}$$

which differs from Equation (1) only by one minus sign and the capitalization of Ψ. Although the transformation (29) is only regular in the region $t_1 < t < t_2$, limited by logarithmic singularities at the zeros of $U(t)$, the metric of Equation (30) can clearly be continued to all values of t, and its analyticity verified by a transformation analogous to Equation (20),

$$Q = W + iX + jY + kZ = e^{t/2} e^{k\phi/2} e^{i\theta/2} e^{k\Psi/2}. \tag{31}$$

Rather than continue with the details of this example, we consider a simpler case based on the 2-dimensional metric

$$ds^2 = 2d\psi dt + t d\psi^2 \tag{32}$$

which can be interpreted (by assigning ψ a period of, say, 2π) as a metric of signature $(-,+)$ on the cylinder $S^1 \times R$. The most significant difference between this example and Taub-NUT space is

that $S^1 \times R$ is *not* simply connected and can be covered by the plane $R \times R$ (assign ψ the range $-\infty$ to $+\infty$) while $S^3 \times R$ *is* simply connected and no range for ψ larger than 0 to 4π is possible there. An inequivalent analytic extension of the region $t > 0$ in the metric (32) is obtained by first making the coordinate transformation

$$\psi = \Psi - 2\ln t \tag{33}$$

on the $t > 0$ region. A simple computation using

$$d\psi = d\Psi - 2t^{-1}dt \tag{34}$$

in Equation (32) gives

$$ds^2 = -2d\Psi dt + td\Psi^2. \tag{35}$$

The coordinate transformation (33) shows the equivalence of the metrics (32) and (35) on the regions $t > 0$. To show their inequivalence when each is considered to define a metric manifold on which t varies from $-\infty$ to $+\infty$ we look at the curve defined on the ψt cylinder by

$$\begin{aligned} \psi &= 0, \\ t &= -\lambda, \end{aligned} \tag{36}$$

with $-\infty < \lambda < +\infty$. Since $\psi(\lambda)$ and $t(\lambda)$ are each analytic functions of λ here this is an analytic curve. Its image on the $t > 0$ part of the Ψt cylinder is also an analytic curve

$$\begin{aligned} \Psi &= 2\ln(-\lambda), \\ t &= -\lambda, \end{aligned} \tag{37}$$

where $\lambda < 0$. However the tangent vector to this curve is

$$\begin{aligned} v^{\Psi} &= d\Psi/d\lambda = 2/\lambda, \\ v^{t} &= dt/d\lambda = -1, \end{aligned} \tag{38}$$

and becomes infinite (i.e., $v^{\Psi} \to -\infty$) as $\lambda \to 0$. Since the analytic structure of the Ψt cylinder is defined by taking $d\Psi$ to be an analytic differential, an analytic extension of this curve segment ($-\infty < \lambda < 0$) would have to make $d\Psi/d\lambda$ an analytic function at $\lambda = 0$, which is impossible. Thus this curve cannot be extended in the Ψt manifold, while it can be in the ψt manifold, demonstrating the inequivalence of these two manifolds.

The geodesics for the metric (32) are easily obtained. One finds that any geodesic segment from a finite point $(\psi_0 t_0)$ to a boundary $(t = \pm \infty)$ is infinitely long in the sense of the affine path parameter, so this space is maximal. However not all geodesics can be extended to parameter values of both $\lambda \to +\infty$ and $\lambda \to -\infty$. For instance

$$\begin{aligned} t &= 0, \\ \psi &= -2\ln\lambda \end{aligned} \tag{39}$$

is a (null) geodesic, where λ has a maximum range $0 < \lambda < \infty$. Consequently this space is not complete.

A further curious property of the example of Equation (32) which was discovered in a discussion with Bonnor is that it is flat. This one may verify by starting from the flat metric

$$ds^2 = d\xi^2 - d\eta^2 = d(\xi + \eta)\, d(\xi - \eta) \tag{40}$$

and making the transformation

$$\begin{aligned} \xi + \eta &= 2te^{\psi/2}, \\ \xi - \eta &= -2e^{-\psi/2}, \end{aligned} \tag{41}$$

under which the half plane $\eta > \xi$ covers the $t\psi$ cylinder infinitely many times.

References

1. L. Auslander and R. E. MacKenzie, *Introduction to differentiable manifolds*, McGraw-Hill, New York, 1963.

2. H. C. Corben and P. Stehle, *Classical mechanics*, 2nd ed., Wiley & Sons, New York, 1960.

3. C. W. Misner, J. Math. Phys. 4 (1963), 924-937.

4. E. Newman, L. Tamburino and T. Unti, J. Math. Phys. 4 (1963), 915-923.

5. A. Taub, Ann. of Math. 53 (1951), 472.

6. A. Taub and C. W. Misner, Ž. Èksper. Teoret. Fiz. 55 (1968), 233-255; English text, Soviet Physics JETP 28 (1969), 122-133.

UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND

A. H. Taub

Relativistic Hydrodynamics

1. **Introduction.** We begin our discussion of relativistic hydrodynamics with the special relativity formulation of the subject. We shall consider a fluid as a collection of a number of particles, each with a rest mass m, which are subject to an external force, and which may interact through collisions. If we write

$$\xi_i = \frac{v_i}{(1 - v^2/c^2)^{1/2}}, \qquad v_i = \frac{\xi_i}{(1 - \xi^2/c^2)^{1/2}}, \qquad i = 1,2,3, \tag{1.1}$$

where

$$v^2 = \sum_{i=1}^{3} v_i^2, \qquad \xi^2 = \sum_{i=1}^{3} \xi_i^2,$$

then, if the v_i are the components of the velocity of a particle relative to an inertial coordinate system in Minkowski space-time, the ξ_i are the components of the momentum per unit rest mass of this particle relative to the same coordinate system. Let $f(t, x^i, \xi_i)$ be the number of particles in the region x^i to $x^i + dx^i$ with values of ξ_i between ξ_i and $\xi_i + d\xi_i$ at time t in the chosen coordinate system. The Boltzman equation for f is then

$$Df = \frac{\partial f}{\partial t} + \frac{\xi_i}{(1 + \xi^2/c^2)^{1/2}} \frac{\partial f}{\partial x^i} + F_i \frac{\partial f}{\partial \xi_i} = \Delta_c(f) \tag{1.2}$$

where F_i is the external force per unit mass and $\Delta_c(f)$ is the time rate of change of f due to collisions between the particles.

Let x stand for the four coordinates x^1, x^2, x^3, and t.

If $f(x,\xi_i)$, $f'(x',\xi_i')$ are the distribution functions of a gas with respect to two different inertial coordinate systems and if x, x' and ξ_i, ξ_i' belong to the same event and the same four-momentum, respectively, then $f(x,\xi_i) = f'(x',\xi_i')$; in this sense the distribution function is a scalar. In fact, the equation

$$\frac{d_3\xi}{(1+\xi^2/c^2)^{1/2}} = \frac{d_3\xi'}{(1+\xi'^2/c^2)^{1/2}}$$

expresses that the parameter-intervals $d_3\xi$, $d_3\xi'$ describe the same infinitesimal domain of four-momenta, and

$$(1+\xi^2/c^2)^{1/2} d_3x = (1+\xi'^2/c^2)^{1/2} d_3x'$$

holds if d_3x, d_3x' are the volumes of two cross-sections $t = \text{const.}$, $t' = \text{const.}$, respectively, of a world tube swept out in space-time by a cloud of particles with four-momenta close to $m(\xi_i, (1+\xi^2/c^2)^{1/2})$ (Lorentz-contraction). Hence, $d_3x d_3\xi = d_3x' d_3\xi'$ holds under conditions where the same particles "belong to" $d_3x\, d_3\xi$ and to $d_3x'\, d_3\xi'$, which proves our assertion, cf. [13].

In terms of the function f we may define the following functions of x_i and t:

$$U^\mu(x) = \int V^\mu(\xi)\, f(x,\xi_i)\, \frac{d_3\xi}{(1+\xi^2/c^2)^{1/2}} = \int V^\mu(\xi)\, d\mu(\xi) \tag{1.3}$$

and

$$T^{\mu\nu}(x) = mc^2 \int V^\mu(\xi)\, V^\nu(\xi)\, f(x,\xi)\, \frac{d_3\xi}{(1+\xi^2/c^2)^{1/2}} = T^{\nu\mu}(x),$$

$$\mu,\nu = 1,2,3,4, \tag{1.4}$$

$$= mc^2 \int V^\mu(\xi)\, V^\nu(\xi)\, d\mu(\xi),$$

where

$$V^i = \xi_i/c, \qquad V^4 = (1+\xi^2/c^2)^{1/2},$$

and hence

$$g_{\mu\nu} V^\mu V^\nu = (V^4)^2 - \sum_{i=1}^{3} (V^i)^2 = 1, \tag{1.5}$$

and the integrations are carried out over the entire volume of ξ_1, ξ_2, ξ_3-space $(-\infty \leqq \xi_i \leqq \infty)$.

Since, as remarked above, $d_3\xi/(1+\xi^2/c^2)^{1/2}$ is Lorentz-invariant, and f is a scalar and $V^\mu(\xi)$ are four-vectors, it follows from these definitions that $U^\mu(x)$ is a four-vector field and that $T^{\mu\nu}(x)$ is a tensor field in space-time.

We next assume that the laws of collision are such that

$$\int \phi^\alpha \Delta_c(f)\, d_3\xi = 0, \qquad \alpha = 0,1,2,3,4, \tag{1.6}$$

where

$$\phi^0 = m, \qquad \phi^i = m\xi_i, \qquad \phi^4 = m(1+\xi^2/c^2)^{1/2}.$$

This assumption is the special relativistic equivalent of the assumption that the total rest mass, momentum and energy of the collection of the particles in unit volume at x is conserved in the collisions.

In case $\alpha = 0$ in Equation (1.6) it follows that we must have

$$m\int Df d_3\xi = m\int \left(\frac{\partial f}{\partial t} + \frac{\xi_i}{(1+\xi^2/c^2)^{1/2}}\frac{\partial f}{\partial x^i} + F_i\frac{\partial f}{\partial \xi_i}\right) d_3\xi = 0,$$

or

$$m(\partial U^4/\partial t + cU^i \partial f/\partial x^i) = mc(U^\alpha)_{,\alpha} = 0, \tag{1.7}$$

where

$$f_{,i} = \frac{\partial f}{\partial x^i}, \qquad f_{,4} = \frac{1}{c}\frac{\partial f}{\partial t}.$$

We shall define the scalar rest mass density of the fluid by the equation

$$\rho^2 = m^2 U^\alpha U_\alpha$$

and write

$$U^\alpha = \frac{\rho}{m}u^\alpha.$$

Then

$$u^\alpha u_\alpha = 1,$$

and Equation (1.7) becomes

$$(\rho u^\mu)_{,\mu} = 0. \tag{1.8}$$

This equation is known as the equation of conservation of mass in classical hydrodynamics. It is often referred to as the equation of conservation of particle number.

In case $\alpha = i$ in Equation (1.6) it follows that we must have

$$\int m\,\xi_i\,Df\,d_3\xi = m\int\left(\xi_i\frac{\partial f}{\partial t} + \frac{\xi_i\xi_j}{(1+\xi^2/c^2)^{1/2}}\frac{\partial f}{\partial x^j} + \xi_i F_j\frac{\partial f}{\partial \xi^j}\right) d_3\xi \tag{1.9}$$

$$= T^{i\mu}{}_{,\mu} - m\,U^4(x)\,F_i = 0.$$

The last term in the last equation arises from the corresponding term in the next to last equation by an integration by parts. Equation (1.9) and the equation arising from Equation (1.6) by setting $\alpha = 4$ may be written as

$$T^{\mu\nu}{}_{,\nu} = \rho\,\mathscr{F}^\mu, \tag{1.10}$$

where

$$\mathscr{F}^i = \frac{F_i}{(1-u^2/c^2)^{1/2}}, \qquad \mathscr{F}^4 = F_i u^i/c$$

is the four-dimensional force vector per unit rest mass.

We shall consider Equations (1.8) and (1.10) as the laws of hydrodynamics in special relativity. In an arbitrary coordinate system in Minkowski space-time and when there are no external forces present they may be written as

$$(\rho u^\mu)_{;\mu} = 0 \tag{1.11}$$

and

$$T^{\mu\nu}{}_{;\nu} = 0. \tag{1.12}$$

There are five equations, corresponding to the five conservation conditions described by Equations (1.6).

2. **Specific internal energy.** The tensor $T^{\mu\nu}$ may be expressed in terms of the time-like unit vector u^μ as

$$T^{\mu\nu} = wu^\mu u^\nu + W^\mu u^\nu + W^\nu u^\mu + W^{\mu\nu}, \tag{2.1}$$

where

$$\begin{aligned} w &= u^\mu u^\nu T_{\mu\nu}, \\ W^\mu &= T_{\rho\nu}u^\nu(g^{\rho\mu} - u^\rho u^\nu), \\ W^{\mu\nu} &= T_{\rho\sigma}(g^{\rho\mu} - u^\rho u^\mu)(g^{\sigma\nu} - u^\sigma u^\nu), \end{aligned} \tag{2.2}$$

and hence

$$W^{\mu}u_{\mu} = W^{\mu\nu}u_{\nu} = 0. \tag{2.3}$$

The vector W^{μ} is known as the heat flow-vector and the tensor $W^{\mu\nu}$ is known as the stress-tensor.

It follows from Equation (2.1) that

$$T = g_{\mu\nu}T^{\mu\nu} = w - 3p, \tag{2.4}$$

where

$$-3p = W^{\mu\nu}g_{\mu\nu}. \tag{2.5}$$

The scalar function p is called the hydrostatic pressure.

We define the rest specific internal energy of the fluid by the equation

$$w = \rho(c^2 + \epsilon).$$

It is then a consequence of Equations (2.2), (1.3), (1.4), (1.8) and Schwarz's inequality that ϵ when considered as a function of ρ and p must satisfy the inequality

$$\epsilon \geqq \frac{3}{2}p/\rho + c^2\left[\left(1 + \frac{9}{4}\left(\frac{p}{\rho c^2}\right)^2\right)^{1/2} - 1\right]. \tag{2.6}$$

The proof of the inequality (2.6) is as follows: Schwarz's inequality states that

$$\left(\int g(\xi)\, d\mu(\xi)\right)^2 \leqq \left(\int g^2(\xi)\, d\mu(\xi)\right)\left(\int d\mu(\xi)\right)$$

for functions $g(\xi)$ for which the integrals exist. Now define $g(\xi)$ by the equation

$$g(\xi) = V^{\mu}(\xi)\, U_{\mu}.$$

Hence

$$\int g(\xi)\, d\mu(\xi) = U^{\mu}U_{\mu} = \frac{\rho^2}{m^2}.$$

Further, it follows from Equations (1.4) and (1.5) that

$$w - 3p = T = mc^2 \int g_{\mu\nu}\, V^{\mu}V^{\nu} d\mu = mc^2 \int d\mu$$

and that

$$T_{\mu\nu} U^\mu U^\nu = mc^2 \int g^2(\xi)\, d\mu(\xi) = \frac{\rho^2}{m^2} w.$$

Hence we must have

$$\rho^2 c^4 \leqq (w - 3p) w,$$

that is, the inequality (2.6) must hold.

In classical hydrodynamics, the specific internal energy ϵ is written as

$$\epsilon = \frac{1}{\gamma - 1} p/\rho. \tag{2.7}$$

If we consider Equation (2.7) as defining γ, no longer a constant, we may write (2.6) as

$$\gamma \leqq 1 + \frac{2}{3} \frac{x}{x + \frac{2}{3}\left(\left(1 + \frac{9}{4} x^2\right)^{1/2} - 1\right)}, \tag{2.8}$$

where

$$x = \frac{p}{\rho c^2} \geqq 0. \tag{2.9}$$

For x small compared to one the right-hand side of the inequality (2.8) is approximately 5/3 whereas for x large compared to one it is approximately 4/3.

3. **Perfect fluids.** Such fluids are those for which the distribution function $f(x, \xi_i)$ gives

$$W^\mu = 0$$

and

$$W^{\mu\nu} = -p(g^{\mu\nu} - u^\mu u^\nu),$$

i.e.,

$$T^{\mu\nu} = \rho(c^2 + i) u^\mu u^\nu - p g^{\mu\nu}. \tag{3.1}$$

Here

$$i = \epsilon + p/\rho \tag{3.2}$$

is called the rest specific enthalpy. It and ϵ are functions of p and ρ.

It follows from Equations (3.1) and (1.9) that

$$T^\mu{}_\nu u^\nu = \rho(c^2 + \epsilon) u^\mu$$

and that

$$T^{\mu}{}_{\nu}X^{\nu} = -pX^{\nu}$$

if

$$u_{\nu}X^{\nu} = 0.$$

That is, the stress energy tensor of a perfect fluid may be characterized algebraically by the following properties: (1) it has a single time-like vector as a proper vector with a positive proper value $\rho(c^2+\epsilon)$, and it has three space-like proper vectors each with the same proper value $-p$. Such stress energy tensors satisfy the equation

$$(T^{\mu}{}_{\nu} + p\delta^{\mu}{}_{\nu})(T^{\nu}{}_{\lambda} - \rho(c^2+\epsilon)\,\delta^{\nu}{}_{\lambda}) = 0, \tag{3.3}$$

where

$$\begin{aligned} p &= -T/4 + (12S - 3T^2)^{1/2}/12, \\ \rho(c^2+\epsilon) &= T/4 + (12S - 3T^2)^{1/2}/4 \\ T &= T^{\mu}{}_{\mu}, \\ S &= T^{\mu}{}_{\rho}T^{\rho}{}_{\mu}, \end{aligned} \tag{3.4}$$

and are characterized by this equation. Hence if we are given $T^{\mu}{}_{\nu}$ we may compute T and S (p and $\rho(c^2+\epsilon)$) and verify if Equation (3.3) is satisfied. If it is we may then be assured that it is the stress energy tensor of a perfect fluid and determine its velocity field by determining the time-like proper vector of $T^{\mu}{}_{\nu}$. From this point of view there is no reason for imposing a restriction such as that given by the inequality (2.6) on ϵ. We shall see, however, that unless an equality such as (2.6) is imposed we will not be able to insure that the velocity of sound in the fluid is less than the velocity of light as it must be in special relativity.

4. **The equations $T^{\mu\nu}{}_{;\nu} = 0$.** Before discussing these equations we introduce the specific entropy S and temperature Θ of the fluid as in classical thermodynamics. We define $S(p,\rho)$ and $\Theta(p,\rho)$ by the requirement that Θ be an integrating factor so that

$$dS = \frac{1}{\Theta}\left(d\epsilon + pd\left(\frac{1}{\rho}\right)\right) \tag{4.1}$$

is a perfect differential. We may also write this equation as

$$\theta\, dS = di - \frac{1}{\rho}\, dp. \tag{4.2}$$

It follows from Equation (3.1) that

$$T^{\mu\nu}{}_{;\nu} = (\rho(c^2 + i)\, u^\nu)_{;\nu} u^\mu + \rho(c^2 + i)\, u^\mu{}_{;\nu} u^\nu - p_{,\nu} g^{\nu\mu}.$$

Hence if the right-hand side of this equation is set equal to zero and the resulting equation is multiplied by u_μ and summed we obtain, taking into account $u_\mu u^\nu = 1$,

$$(\rho(c^2 + i)\, u^\nu)_{;\nu} - p_{,\nu} u^\nu = 0 \tag{4.3}$$

and

$$\rho(c^2 + i)\, u^\mu{}_{;\nu} u^\nu = p_{,\nu}(g^{\mu\nu} - u^\mu u^\nu) \tag{4.4}$$

as equations equivalent to $T^{\mu\nu}{}_{;\nu} = 0$. Equations (4.4) are the relativistic analogues to the equations of conservation of momentum. We shall now see that Equations (4.3) alone do not correspond to a classical conservation theorem. This equation may be written as

$$(c^2 + i)(\rho u^\nu)_{;\nu} + \rho \left(\epsilon_{,\mu} + p \left(\frac{1}{\rho} \right)_{,\mu} \right) u^\mu = 0$$

or as

$$(c^2 + i)(\rho u^\nu)_{;\nu} + \rho\theta S_{,\mu} u^\mu = 0. \tag{4.5}$$

In classical hydrodynamics, it can be shown that when shock waves are absent then the equation describing the conservation of energy and the equations of conservation of momentum imply that the entropy is conserved along the stream-lines, that is along the world lines of the elements of the fluid. The relativistic form of this statement is

$$S_{,\mu} u^\mu = 0. \tag{4.6}$$

Equation (4.6) follows from the conservation law $T^{\mu\nu}{}_{;\nu} = 0$ if and only if the law of conservation of mass, the equation

$$(\rho u^\mu)_{;\mu} = 0, \tag{4.7}$$

also holds. Equation (4.6) does not imply that $S =$ constant. If however S is constant on some initial hypersurface and if no shocks are present it will be constant throughout the fluid.

We may write Equation (4.3) as

$$\rho(c^2+i)\,u^{\nu}{}_{;\nu}+[\rho(c^2+\epsilon)]_{,\nu}u^{\nu}=0. \tag{4.8}$$

If the motion of the fluid is such that $\rho(c^2+\epsilon)$ is a function of $\rho(c^2+i)=\rho(c^2+\epsilon)+p$ alone, then we may define a quantity α by the equation

$$d\alpha/\alpha = d[\rho(c^2+\epsilon)]/(\rho(c^2+\epsilon)+p),$$

and Equation (4.8) becomes

$$(\alpha u^{\mu})_{;\mu}=0.$$

We shall have

$$\alpha=\rho$$

if and only if

$$dS=0.$$

5. **The Rankine-Hugoniot equations.** From a study of Equations (1.11) and (1.12) in Minkowski space-time for the case of one-dimensional motion of a perfect fluid (where $u^2=u^3=0$ and all quantities are functions of x^1 and x^4 alone) it can be shown [1] that the velocity of sound is given by

$$a^2=\frac{\rho}{(1+i/c^2)}\left(\frac{di}{d\rho}\right)_S \tag{5.1}$$

where the entropy S is kept constant in evaluating $di/d\rho$.

In case the enthalpy is a function of $x=p/\rho c^2$ alone, we may write Equation (5.1) as

$$a^2=i'x/(i'-c^2)(1+i/c^2)$$

where

$$i'=di/dx$$

and

$$\frac{\rho di}{c^2 d\rho}=\frac{\rho i'}{c^2}\frac{dx}{d\rho}=\rho\frac{dx}{d\rho}+x;$$

the last equation is a consequence of the fact that $dS/d\rho=0$.

A fluid for which the equality holds in (2.6) will be called a limiting fluid. For such a fluid

$$1+i/c^2=5x/2+(1+9x^2/4)^{1/2}$$

and

$$\lim_{x\to\infty} a^2 = \frac{1}{3}, \tag{5.2}$$

$$\lim_{x\to\infty} \rho(c^2+\epsilon) = 3p. \tag{5.3}$$

This limit for the velocity of sound in an extreme relativistic gas was obtained by Curtis [2] and de Hoffman and Teller [3]. Taub [1] and Israel [4] have shown by considering sound waves as infinitesimal shock waves the relation between the requirement that the speed of sound be less than the speed of light and certain conditions that must obtain between the various thermodynamic variables that characterize the fluid.

The discussion of one-dimensional motion of a fluid in Minkowski space given in [1] shows that shocks form from compressive motions just as in the classical theory. When this occurs, Equations (1.11) and (1.12) cannot be applied at the shocks and must be replaced by statements that relate the variables describing the flow of matter, momentum and energy across the shocks. These are the relativistic Rankine-Hugoniot equations. They are derived as follows: Equations (1.11) and (1.12) are equivalent to the statements that

$$(\rho u^\mu f)_{;\mu} = \rho u^\mu f_{,\mu} \tag{5.4}$$

and

$$(T^{\mu\nu}\lambda_\nu)_{;\mu} = T^{\mu\nu}\lambda_{\mu;\nu} \tag{5.5}$$

for arbitrary functions f and vector fields λ_μ which have continuous first derivatives.

In view of Gauss' theorem we may write these equations as

$$\int \rho u^\mu f n_\mu dv = \int \rho u^\mu f_{,\mu}(-g)^{1/2} d^4x \tag{5.6}$$

and

$$\int T^{\mu\nu}\lambda_\nu n_\mu dv = \int T^{\mu\nu}\lambda_{\mu;\nu}(-g)^{1/2} d^4x. \tag{5.7}$$

The integrals on the right in Equations (5.6) and (5.7) refer to an arbitrary four dimensional volume and those on the left refer

to a closed hypersurface enclosing this volume. Equations (5.6) and (5.7) are meaningful even when the integrands are discontinuous. We assume they hold in case ρ, u^μ and $T^{\mu\nu}$ are discontinuous across a hypersurface Σ in Minkowski space which is the history of a two-dimensional surface in the spaces $t =$ constant.

By enclosing an arbitrary portion Σ' of the hypersurface in a thin four-dimensional volume and taking the limit as this volume shrinks to zero we may show that

$$\int_{\Sigma'} [\rho u^\mu] f n_\mu dv = 0 \tag{5.8}$$

and

$$\int_{\Sigma'} [T^{\mu\nu}] \lambda_\nu n_\mu dv = 0, \tag{5.9}$$

where we have used the notation

$$[F] = F_+ - F_-$$

where F_+, F_- are the boundary values of F on the two sides of Σ.

Since the integrals in Equations (5.8) and (5.9) refer to arbitrary portions of Σ and since f and λ_μ are arbitrary, we must have

$$[\rho u^\mu] n_\mu = 0, \tag{5.10}$$

$$[T^{\mu\nu}] n_\mu = 0. \tag{5.11}$$

In case the $T^{\mu\nu}$ entering into Equations (5.11) is given by Equations (8.1), Equations (5.10) and (5.11) are called the relativistic Rankine-Hugoniot equations, the equations that express the conservation of mass, energy and momentum across a singular hypersurface in space-time, a hypersurface on which the flow variables may be discontinuous. The derivation given above holds in a general space-time.

If in Minkowski space-time in an inertial coordinate system the parametric equations of Σ are given by

$$x^\mu = X^\mu(u, V, t), \tag{5.12}$$

that is,

$$\begin{aligned} t &= t, \\ x^i &= X^i(u, V, t), \end{aligned} \tag{5.13}$$

then the three-dimensional velocity of a point labelled by u, V on the hypersurface is given by

$$V^i = \partial X^i / \partial t,$$

and the three-dimensional normal to the surfaces in three-space given by the last three of Equations (5.13) has components proportional to

$$\Lambda_i = \epsilon_{ijk} \frac{\partial X^j}{\partial u} \frac{\partial X^k}{\partial V}.$$

We normalize the Λ_i by requiring

$$\sum_{i=1}^{3} \Lambda_i^2 = 1.$$

Then the quantity

$$V = \Lambda_i V^i$$

is the velocity of these surfaces in three-space in the direction of the normal. The four-vector

$$n_\mu = \left(\Lambda_1, \Lambda_2, \Lambda_3, \frac{V}{c}\right)$$

is normal to the hypersurface Σ. We may verify that

$$n_\mu n^\mu = \frac{V^2}{c^2} - \sum_i \Lambda_i^2 \leqq (\sum \Lambda_i^2) \left(\frac{\sum V_i^2}{c^2} - 1\right),$$

and hence

$$n_\mu n^\mu \leqq 0$$

is equivalent to

$$\sum_{i=1}^{3} V_i^2 \leqq c^2. \tag{5.14}$$

We define a unit normal vector

$$N_\mu = \alpha n_\mu$$

such that

$$N_\mu N^\mu = -1$$

when Equation (5.14) holds. Note that if u^μ is the four-velocity

of the fluid then

$$u^\mu = (1 - u^2/c^2)^{-1/2}(u_1/c, u_2/c, u_3/c, 1)$$

where u_i are the components of the three-dimensional velocity of the fluid and

$$u^\mu N_\mu = (1 - V^2/c^2)^{-1/2}(1 - u^2/c^2)^{-1/2}(V - U)/c$$

where

$$U = \sum \Lambda_i u_i$$

is the three-dimensional velocity of the fluid in the direction of the normal to the surface. Thus $u^\mu N_\mu$ is proportional to the normal component of the velocity of the surface relative to the fluid.

When Equations (3.1) hold, Equations (5.10) and (5.11) may be written as

$$M = \rho_+ u_+^\mu N_\mu = \rho_- u_-^\mu N_\mu \tag{5.15}$$

and

$$M(\mu_+ u_+^\mu - \mu_- u_-^\mu) = (p_+ - p_-) N^\mu \tag{5.16}$$

with

$$\mu = c^2 + i = c^2 + \epsilon + p/\rho \tag{5.17}$$

respectively. If V_i are the components of a unit three-vector in the surface, then the four-vector

$$Y^\mu = (V_1, V_2, V_3, 0)$$

satisfies the relation

$$Y^\mu N_\mu = 0,$$

and

$$Y^\mu u_\mu = (1 - u^2/c^2)^{-1/2}(V_i u_i/c)$$

is proportional to the velocity of the fluid in the direction V_i. Multiplying Equations (5.16) by Y_μ and summing, we obtain

$$M(\mu_+ u_+^\mu Y_\mu - \mu_- u_-^\mu Y_\mu) = 0. \tag{5.18}$$

Hence, either

$$M = 0, \tag{5.19}$$

or

(5.20) $$\mu_+ u_+^\mu Y_\mu = \mu_- u_-^\mu Y_\mu,$$

or both (5.19) and (5.20) hold. We shall say that the case $M = 0$ represents a slip-stream discontinuity or a density discontinuity and say that the case $M \neq 0$ represents a shock wave. In the former case it follows from Equations (5.15) and the interpretation of $u^\mu N_\mu$ that no matter crosses the surface of discontinuity. That is, this surface is made up of stream-lines of the fluid. Equation (5.20) need not hold when $M = 0$. If it does in this case, we shall say that there is a density discontinuity present.

If ρ_- and $T_-^{\mu\nu}$ both vanish, we are dealing with a particular case for which $M = 0$. And then Equations (5.16) and (5.15) become

$$p_+ = 0$$

and

$$\rho_+ u_+^\mu N_\mu = 0.$$

It is evident that ρ_+ need not vanish if

$$u_+^\mu N_\mu = 0,$$

that is, if the stream-lines of the flow do not cross the surface of discontinuity.

6. **Shock waves.** As was stated earlier for such waves we have $M \neq 0$. It is convenient to rewrite Equations (5.15) and (5.16) in terms of the vector

(6.1) $$V_\lambda = (\mu/c^2)\, u_\lambda$$

and the quantity

(6.2) $$\tau = \rho/\mu.$$

Then

(6.3) $$V^\mu V_\mu = \mu^2/c^4 = (1 + i/c^2)^2,$$

and we have instead of Equations (5.15) and (5.16) the equations

(6.4) $$M = c^2 \tau_+ V_+^\mu N_\mu = c^2 \tau_- V_-^\mu N_\mu,$$

(6.5) $$M(V_+^\mu - V_-^\mu) = (p_+ - p_-)\, N^\mu.$$

Equations (5.20), which must hold for shock waves, become

(6.6) $$V_+^\mu Y_\mu = V_-^\mu Y_\mu$$

for any Y_μ satisfying

$$Y^\mu N_\mu = 0,$$

$$Y^\mu Y_\mu = -1.$$

Equations (6.6) are two of the four equations contained in Equations (6.5). The remaining two will be derived below.

Multiply Equations (6.5) by $V_{+\mu}$ and by $V_{-\mu}$ and sum and obtain

$$\left(\frac{\mu_+^2}{c^4} - V_-{}^\mu V_{+\mu}\right) = (p_+ - p_-)\frac{1}{c^2\tau_+},$$

$$\left(V_+{}^\mu V_{-\mu} - \frac{\mu_-^2}{c^4}\right) = (p_+ - p_-)\frac{1}{c^2\tau_-}.$$

Hence

$$\frac{\mu_+^2}{c^4} - \frac{\mu_-^2}{c^4} = \frac{1}{c^2}(p_+ - p_-)\left(\frac{1}{\tau_+} + \frac{1}{\tau_-}\right). \tag{6.7}$$

When ϵ is a known function of p and ρ, i is determined as a function of p and ρ and so is τ. Then for fixed p_-, ρ_-, the p_+, ρ_+ satisfying Equation (6.7) lie on a curve. This curve connects those (p,ρ)-states that can be obtained from (p_-,ρ_-) by passing through a shock. The curve is known as the Hugoniot curve. We must select that part of the curve passing through p_-, ρ_- corresponding to states p_+, ρ_+ whose entropy $S_+ > S_-$. Otherwise, the second law of thermodynamics would be violated.

Multiply Equations (6.5) by N_μ and sum. We obtain

$$\frac{m^2}{c^2}\left(\frac{1}{\tau_+} - \frac{1}{\tau_-}\right) = -(p_+ - p_-)(-N_\mu N^\mu) \tag{6.8}$$

$$= -(p_+ - p_-).$$

Equations (6.4), (6.6), (6.7), and (6.8) are equivalent to Equations (6.4) and (6.5).

Since shock waves must travel with a velocity less than that of light and since M must be real, we must have

$$\frac{p_+ - p_-}{1/\tau_- - 1/\tau_+} > 0,$$

that is, when

$$p_+ > p_-,$$

we must have

$$\tau_+ > \tau_-.$$

In other words an increase (decrease) in pressure across a shock wave must correspond to an increase (decrease) in τ.

Israel has shown [4] that if $\epsilon(p,\rho)$ is such that

$$(\partial p/\partial \rho)_S > 0,$$

$$(\partial S/\partial p)_\rho > 0,$$

$$(\partial p/\partial \rho)_S < (1 + i/c^2),$$

then the speed of shock waves is always less than the speed of light. The last of these conditions is the requirement that the speed of sound is less than the speed of light, and we have seen that it is satisfied for the limiting fluid as are the first two conditions.

7. **Lagrangian and co-moving coordinates.** Classical hydrodynamics deals with a fluid moving in three-space in which a coordinate system is introduced in which every point of the three-space is labelled by coordinates x^i. Relative to this coordinate system, the velocity vector of the fluid is described by its components $v^i(x,t)$. The v^i are called the Eulerian velocities. The stream-lines of the fluid are then the curves given by the solutions of the ordinary differential equations

$$dx^i/dt = v^i(x,t), \qquad i = 1,2,3. \tag{7.1}$$

These solutions may be written as

$$x^i = x^i(\xi,t), \tag{7.2}$$

where the ξ^i are the initial values of the x^i. That is,

$$x^i = x^i(\xi,0) = \xi^i. \tag{7.3}$$

The ξ^i are called the Lagrange coordinates and may be used to describe the motion of the fluid. Equations (7.2) supplemented by the equation $t = \tau$ may be regarded as a set of equations describing transformation between the variables t, x^i and τ, ξ^i, and the equations of conservation of mass, momentum, and energy may be rewritten in terms of the variables τ, ξ^i.

The congruence of curves given by Equation (7.2) may be pictured in space-time. In this interpretation, they represent the world lines of an element of the fluid which is labelled by the

three parameters ξ^i. These world lines may be described in terms of their four-dimensional tangent vector

$$dx^\mu/dt = w^\mu(x), \qquad \mu = 1,2,3,4, \tag{7.4}$$

where

$$w^i = v^i, \qquad v^4 = c,$$

since

$$x^4 = ct.$$

We define a hypersurface Σ in space-time by the equation

$$t = 0. \tag{7.5}$$

In this hypersurface we label points by the coordinates ξ^i. That is, we describe the hypersurface Σ by the parametric equations

$$x^\mu = X^\mu(\xi^1, \xi^2, \xi^3) \tag{7.6}$$

instead of the functional equation

$$F(x^\mu) \equiv t = 0.$$

Then the congruence of world lines given by Equations (7.2) may be described as the set of solutions of Equations (7.4) which intersect the hypersurface Σ in the points ξ^i of Σ.

Equations (7.4) may be written as

$$dx^\mu/dt = \rho u^\mu,$$

where u^μ is the unit four-dimensional velocity vector of the fluid and ρ is given by the equation

$$\rho^2 = \frac{ds^2}{dt^2} = g_{\mu\nu}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt} = g_{\mu\nu}w^\mu w^\nu.$$

If the parameter which labels points on the world line determined by ξ^i is changed to any other parameter, say ω, Equations (7.4) may still be written as

$$dx^\mu/d\omega = \rho u^\mu = \bar{w}^\mu, \tag{7.7}$$

where now ω is related to ρ by a similar equation to the one given above. If ω is taken to be equal to s, the proper time along the world line, we have

$$dx^\mu/ds = u^\mu.$$

Now let Σ be an arbitrary space-like hypersurface in space-time described by the parametric Equations (7.6), and let us write the solutions of Equations (7.7) as

$$x^\mu = x^\mu(\eta; \omega),$$

where

$$x^\mu = x^\mu(\eta; 0) = \eta^\mu.$$

That is, the η^μ are the initial values of the x^μ. We further require that the η^μ lie on the hypersurface Σ. Then

$$\eta^\mu \equiv X^\mu(\xi),$$

and we may write the solutions of Equations (7.7) as

$$x^\mu = x^\mu(\xi; \omega). \tag{7.8}$$

These equations may be regarded as a transformation of co-ordinates from the coordinates x^μ to the coordinates ξ^1, ξ^2, ξ^3, and ω. The latter coordinates are called co-moving coordinates in space-time. They differ from Lagrangian coordinates in that the hypersurface Σ used in their definition need not be given by the Equation (7.5) in the Eulerian coordinate system.

The functions x^μ and $\bar{w}^\mu$ entering into Equations (7.7) may be regarded as functions of ξ and ω in view of Equations (7.8). If they are so regarded, then they should be written as

$$\partial x^\mu / \partial \omega = \bar{w}^\mu(\xi, \omega), \tag{7.9}$$

and it should be understood that the ξ^i are kept constant in the differentiation involved on the left-hand side of these equations.

From the transformation law of vectors, it follows that in the co-moving coordinate system, the components of the vector $\bar{w}^\mu$ are given by

$$\bar{w}^{\mu *} = \bar{w}^\rho \, \partial x^{\mu *} / \partial x^\rho,$$

where

$$x^{i *} = \xi^i, \qquad x^{4*} = \omega.$$

In view of Equation (7.9), we have

$$\bar{w}^{\mu *} = \frac{\partial x^{\mu *}}{\partial x^\rho} \frac{\partial x^\rho}{\partial w} = \frac{\partial x^{\mu *}}{\partial x^{4*}} = \delta^\mu_4.$$

Hence, in a co-moving coordinate system, the four-velocity vector

of the fluid satisfies the equations

$$u^\mu = \delta^\mu_4 g_{44}^{-1/2} \tag{7.10}$$

since we must have

$$g_{\mu\nu} u^\mu u^\nu = 1.$$

8. **The equations of hydrodynamics in co-moving coordinates.** Co-moving coordinates may be introduced both in special relativity and in general relativity. In the former theory, they complicate the metric in that we no longer have

$$g_{\mu\nu} = \eta_{\mu\nu}.$$

However, Equations (4.4), (4.6), and (4.7) have relatively simple forms in co-moving coordinates.

Thus if Equation (7.10) holds, Equation (4.6) becomes

$$S_{,4} = 0,$$

that is

$$S = S(\xi). \tag{8.1}$$

Equation (4.7) may be written as

$$(g^{1/2} g_{44}^{-1/2} \rho)_{,4} = 0$$

or as

$$\rho = g_{44}^{1/2} (-g)^{-1/2} \rho_0(\xi), \tag{8.2}$$

where g is the determinant of the $g_{\mu\nu}$ and

$$g_{\mu\nu} = \eta_{\alpha\beta} \frac{\partial x^\alpha}{\partial x^{*\mu}} \frac{\partial x^\beta}{\partial x^{*\nu}}, \tag{8.3}$$

$$ds^2 = g_{44} d\omega^2 + 2 g_{4i} d\omega d\xi^i + g_{ij} d\xi^i d\xi^j$$

in the case of special relativity. They are such that the Riemann-Christoffel tensor vanishes:

$$R^\mu{}_{\nu\sigma\tau} = 0. \tag{8.4}$$

Since

$$u^\mu{}_{;\nu} = \frac{\partial u^\mu}{\partial x^{*\nu}} + u^\sigma \Gamma_\sigma{}^\mu{}_\nu$$

where the $\Gamma_\sigma{}^\mu{}_\rho$ are the Christoffel symbols evaluated in this co-

ordinate system, we have in this coordinate system:

$$u^{\mu}_{;\nu} = \frac{\partial}{\partial x^{*\nu}}\left(\frac{1}{g_{44}^{1/2}}\right)\delta_4^{\mu} + \frac{1}{g_{44}^{1/2}}\Gamma_{4\nu}^{\mu},$$

and

$$u^{\nu}u^{\mu}_{;\nu} = \frac{1}{g_{44}^{1/2}}\frac{\partial}{\partial\omega}\left(\frac{1}{g_{44}^{1/2}}\right)\delta_4^{\mu} + \frac{1}{g_{44}}\Gamma_{44}^{\mu},$$

where

$$\Gamma_{44}^{\mu} = g^{\mu\nu}\left(\frac{\partial g_{4\nu}}{\partial\omega} - \frac{1}{2}\frac{\partial g_{44}}{\partial x^{*\nu}}\right).$$

If we define

$$F = \log g_{44}^{1/2}, \tag{8.5}$$

we may write

$$u^{\nu}u^{\mu}_{;\nu} = \left(g^{\mu\nu} - \frac{\delta_4^{\mu}\delta_4^{\nu}}{g_{44}}\right)\frac{\partial F}{\partial x^{*\nu}} + g^{\mu\nu}\frac{\partial}{\partial\omega}\left(\frac{g_{\nu 4}}{g_{44}}\right).$$

Equations (4.4) then become

$$\left(g^{\mu\nu} - \frac{\delta_4^{\mu}\delta_4^{\nu}}{g_{44}}\right)\left(\frac{\partial p}{\partial x^{*\nu}} + \rho c^2\left(1 + \frac{i}{c^2}\right)\frac{\partial F}{\partial x^{*\nu}}\right) = g^{\mu\nu}\frac{\partial}{\partial\omega}\left(\frac{g_{4\nu}}{g_{44}}\right). \tag{8.6}$$

It follows from this equation that if the flow is stationary in co-moving coordinates, that is if all quantities are independent of ω, the time variable in co-moving coordinates, then we must have[†]

$$\frac{\partial p}{\partial\xi^i} + \rho c^2\left(1 + \frac{i}{c^2}\right)\frac{\partial F}{\partial\xi^i} = 0. \tag{8.7}$$

If we recall Equation (8.5) and interpret g_{44} in terms of the Newtonian potential, then Equation (8.7) is recognized as the equation for hydrostatic equilibrium.

Equations (8.7) must hold whenever the right-hand sides of Equations (8.6) vanish. In particular, they must hold when we have a co-moving coordinate system in which $g_{4i} = 0$. It can be shown that this implies that the flow is irrotational.

Time does not permit a further discussion of the equations of hydrodynamics. I shall close this section by referring you to the

[†] This result was communicated to me verbally by Dr. J. R. Wright.

discussion of circulation, Bernoulli's theorem and other extensions of classical results that may be found in reference [5].

9. **General relativistic hydrodynamics.** Problems in this area are of two sorts: (I) Those in which the gravitational field of the fluid can be neglected, but the gravitational field due to some other matter or energy must be taken into account, and (II) those in which the only matter and gravitational fields which are present are the fluid and its own gravitational field.

For problems of type I, the basic equations are Equations (1.11), (1.12), (5.10) and (5.11), but now the covariant differentiation is done with respect to a metric tensor $g^{\mu\nu}$ which is determined by the stress-energy tensor $\theta^{\mu\nu}$, the source of the exterior gravitational field and its potentials, through the equations

$$R^{\mu\nu} - \frac{1}{2} g^{\mu\nu} R = - kc^2 \theta^{\mu\nu}, \tag{9.1}$$

where $R^{\mu\nu}$ is the Ricci curvature tensor computed from the $g^{\mu\nu}$, R is the scalar curvature, and

$$k = 8\pi G/c^2$$

with G the Newtonian constant of gravitation. Thus for this type of problem the only difference between the special theory of relativity and the general theory is that the underlying space-time is changed from a Minkowski one to a general one whose metric tensor satisfies Equation (9.1).

In problems of type II Equations (9.1) with $\theta^{\mu\nu} = T^{\mu\nu}$ hold. Thus we must determine both $T^{\mu\nu}$ and $g^{\mu\nu}$ so that

$$R^{\mu\nu} - \frac{1}{2} g^{\mu\nu} R = - kc^2 T^{\mu\nu} \tag{9.2}$$

and Equations (3.1), (1.11), (1.12), (5.10) and (5.11) obtain. Equations (1.12) are a consequence of Equations (9.2). The former may be considered as a first integral of equations (9.2), and this observation has led McVittie [6] to a method for solving problems in prerelativity hydrodynamics.

Equations (3.1), (1.8), and (9.2) are sometimes said to describe the interior problem for the Einstein field equations. The "solution" of this problem requires the determination of the ten components of the tensor $g^{\mu\nu}$, three independent components of the four-velocity

vector u^μ and the two scalar functions p and ρ. The scalar function ϵ is supposed to be known as a function of p and ρ when the nature of the fluid medium is given. Thus one is required to "solve" the ten Equations (9.2) when neither the right-hand side nor the left-hand side of these equations is given. Surprising as it may seem this can be done at least in a variety of cases. In the following we shall see how this can be done by using the algebraic properties of the $T^{\mu\nu}$ satisfying Equations (3.1) or by imposing the conservation of mass (Equation (1.11)) condition and the condition of isentropy and irrotationality.

By substituting for $T^\mu{}_\nu$ from Equations (9.2), Equations (3.3) and (3.4) become a set of partial differential equations for the $g_{\mu\nu}$ alone. The solutions of these equations may be substituted into Equations (9.2) to determine $T^\mu{}_\nu$. The equations for the $g_{\mu\nu}$ obtained by the method described above have been called the consistency equations by McVittie [6]. He dealt with an approximate form of these equations and thus obtained various solutions for problems in pre-relativity hydrodynamics. Rainich [7] and Misner and Wheeler [8] used a similar technique for formulating the Einstein-Maxwell field equations for the case where the only stress-energy tensor creating the gravitational field is that due to the electromagnetic field and no electromagnetic sources are present.

The vorticity vector v^μ is defined by the equation

$$v^\mu = (-g)^{-1/2}\epsilon^{\mu\nu\sigma\tau}u_{\sigma;\tau}u_\nu, \tag{9.3}$$

where g is the determinant of the metric tensor and $\epsilon^{\mu\nu\sigma\tau}$ is the alternating tensor density whose components are equal to plus or minus one depending on whether $(\mu\nu\sigma\tau)$ is an even or odd permutation of (1234). A flow will be said to be irrotational if

$$v^\mu = 0.$$

If we define the vector

$$V^\mu = (1 + i/c^2)\, u^\mu,$$

it may be shown [5] as a consequence of Equations (2.6) that

$$\Omega_{\mu\nu} = (1 + i/c^2)(-g)^{1/2}v^\sigma u^\tau\epsilon_{\sigma\tau\mu\nu} + TS_{;\nu}u_\mu - TS_{;\mu}u_\nu = V_{\mu,\nu} - V_{\nu,\mu}$$

where T is the temperature and S is the entropy.

Hence for isentropic (S = constant), irrotational ($v^\mu = 0$) flow we have

$$\Omega_{\mu\nu} = 0.$$

As a consequence

$$u_\mu = \frac{1}{1 + i/c^2} \frac{\partial \theta}{\partial x^\mu},$$

where θ is arbitrary scalar function. We may adapt our coordinate system so that $\theta = x^4$, and then in this comoving coordinate system

$$u_\mu = \frac{1}{(1 + i/c^2)} \delta^4{}_\mu.$$

It then follows from Equation (1.9) that

$$g^{44} = (1 + i/c^2)^2.$$

Hence for irrotational isentropic flow we may choose a comoving coordinate system in which the stress-energy tensor depends on g^{44} alone, since in this case the enthalpy, i, is a function of pressure alone as is the rest density. If these functions of g^{44} and the above expression for u_μ are used in the evaluation of the tensor $T^{\mu\nu}$ we may write the field Equations (9.2) wholly in terms of the tensor $g_{\mu\nu}$. Thus the Einstein equations for the case of relativistic hydrodynamics involve a single set of dependent variables and one has a not impossible problem.

Various exact and approximate solutions of these equations have been discussed. In particular, the problem of the one-dimensional motion of a gas analogous to the problem in classical hydrodynamics of the motion generated by pushing a piston into a tube filled with gas has been treated approximately [9], [10]. The procedure starts with the special relativistic solution to the problem and modifies this solution in accordance with the Einstein field equations. It can be seen that the nonlinearities in the problem reside in the special relativistic approximations and that the gravitational corrections satisfy linear equations. Thus, in the zero approximation shocks occur. It is unlikely that the gravitational corrections satisfying linear equations will remove them. We must therefore be prepared in general relativity to treat with

and allow for coordinate transformations involving discontinuous derivatives. The metric tensor thus becomes singular [11] on certain hypersurfaces. Although such singularities are associated with coordinate transformations they are essential ones since the coordinates involved have intrinsic and physical significance.

References

1. A. H. Taub, *Relativistic Rankine-Hugoniot equations*, Phys. Rev. 74 (1948), 328-334.

2. A. R. Curtis, Proc. Roy. Soc. Ser. A **200** (1950), 248.

3. F. deHoffmann and E. Teller, Phys. Rev. **80** (1950), 692.

4. W. Israel, Proc. Roy. Soc. Ser. A **259** (1960), 129.

5. A. H. Taub, *On circulation in relativistic hydrodynamics*, Arch. Rational Mech. Anal. 3 (1959), 312-324.

6. G. C. McVittie, *General relativity and cosmology*, Chapters VI and VII, Wiley, New York, 1956.

7. G. Y. Rainich, *Electrodynamics in general relativity theory*, Trans. Amer. Math. Soc. **27** (1925), 106-136.

8. C. W. Misner and J. A. Wheeler, *Geometrodynamics*, Ann. Physics 2 (1957), 525-603.

9. ———, *Isentropic hydrodynamics in plane symmetric space-time*, Phys. Rev. **103** (1956), 454-467.

10. ———, *Approximate solutions of the Einstein equations for isentropic motions of plane symmetric distributions of perfect fluids*, Phys. Rev. **107** (1957), 884-900.

11. ———, *Singular hypersurfaces in general relativity*. III, J. Math. 1 (1957), 370-388.

General References

12. A. Lichnerowicz, *Theories relativistes de la gravitation et de l'ectromagnetisme*, Masson, Paris, 1955.

13. J. L. Synge, *The relativistic gas*, North-Holland, Amsterdam, 1957.

14. ———, *Relativity, the general theory*, North-Holland, Amsterdam, 1960.

UNIVERSITY OF CALIFORNIA, BERKELEY
BERKELEY, CALIFORNIA

G. C. McVittie

Gravitational Collapse

The notion of gravitational collapse has come into prominence recently because it seems to provide a simple way of concentrating very large amounts of energy into small volumes. Hoyle and Fowler [1] have pointed out that the gravitational potential energy, calculated by Newtonian theory, of a spherical mass M of radius R is of the order of GM^2/R. Thus a mass of the order of $10^8 M_\odot$, concentrated into a sphere of radius $3.8 \times 10^4 R_\odot$, would have a gravitational potential energy of some 10^{60}ergs. Such energies are required to maintain the optical and radio emissions of quasi-stellar radio sources for a reasonable length of time. These objects also apparently occupy surprisingly small volumes of space.

I shall try to describe, on the basis of general relativity, some of the methods that have been employed to analyze the motions of collapse from configurations of very large radius to extremely small ones.* I do not attempt to suggest how these motions originate, from the breakdown of a previous state of equilibrium, for example. The initial conditions will simply be that the configuration is either initially at rest, or in a state of radial motion, and that this initial

*A more detailed review will be found in G. C. McVittie, Astrophys. J. 140 (1964), 401.

state is not one of equilibrium under the action of gravitational self-attraction and, if it is present, of a pressure-gradient.

The simplest theory that has been employed is that of the expanding universe in reverse—with a velocity of contraction instead of expansion—combined with the requirements that the mass shall be constant and finite and that the outer radius of the (spherical) configuration be finite also. Inside the collapsing spherical mass the metric is

$$ds^2 = dt^2 - \frac{R^2(t)}{c^2 R_0^2}\left\{\frac{dr^2}{1 - kr^2/R_0^2} + r^2 d\Omega^2\right\}, \tag{1}$$

where $R_0 = R(0)$ and $t = 0$ is regarded as the initial instant. The space-curvature constant k has one of the customary three values $+1, 0$ or -1. Outside the material there is a Schwarzschild space-time with metric

$$ds^2 = \left(1 - \frac{2m}{u}\right) d\bar{t}^2 - \frac{1}{c^2}\left\{\frac{du^2}{1 - 2m/u} + u^2 d\Omega^2\right\}, \tag{2}$$

where, in both cases,

$$d\Omega^2 = d\Theta^2 + \sin^2\Theta d\phi^2. \tag{3}$$

In the field (1) write

$$\tau = t/T, \qquad S(\tau) = R/R_0, \tag{4}$$

where T is a constant to be determined later. The energy-tensor in this field reduces to a density ρ and an isotropic pressure p, which, by Einstein's equations, are given by

$$\frac{8\pi G T^2}{3}\rho = k\frac{c^2 T^2}{R_0^2}\frac{1}{S^2} + \frac{S'^2}{S^2}, \tag{5}$$

$$8\pi G T^2 \frac{p}{c^2} = -\frac{2S''}{S} - \frac{S'^2}{S^2} - k\frac{c^2 T^2}{R_0^2}\frac{1}{S^2}, \tag{6}$$

where a prime denotes a derivative with respect to τ. These two equations show that, if the metric is postulated to have the form (1), then the density and pressure are necessarily functions of the time alone and so are spatially uniform at each instant of time. The Equations (5) and (6) may be solved by assuming that there is, for example, a polytropic relation between p and ρ. But this would mean that the pressure would have to be nonzero at the outer

boundary of the configuration. The field (1) cannot then be fitted onto the vacuum field (2). Therefore the approach via an equation of state is impossible when (1) is taken as the starting point, except in one case. This possibility is that the pressure is always zero within, and on the boundary of, the configuration. The consequences of this zero-pressure hypothesis will now be examined.

Zero-pressure collapse. The initial density is ρ_0, the initial value of S is unity, and the initial rate of change of S is S_0'. Write

$$\alpha = 8\pi G\rho_0 T^2/3, \qquad \beta = (-S_0'), \tag{7}$$

and then (5) and (6), with $p = 0$, evaluated at the initial instant lead to

$$\alpha - \beta^2 = kc^2T^2/R_0^2, \tag{8}$$

$$S_0'' = -\alpha/2, \quad (<0). \tag{9}$$

The Equation (5) can then be replaced by

$$S'^2 = \alpha/S - (\alpha - \beta^2). \tag{10}$$

Since $\alpha > 0$, it follows from (9) that collapse always ensues; it also follows from (8) that the collapse can start from rest ($\beta = 0$) only if $k = +1$, which means that space is closed. When $k = 0$ or $k = -1$, the motion must necessarily start with $\beta \neq 0$, and the material must thus initially be in a state of radial motion. Obviously such an initial state of radial motion is also compatible with $k = +1$ provided that $\beta^2 < \alpha$.

The Equation (10) may be integrated in terms of elementary functions for all three values of k. The constant T is then interpretable as the time required for the collapse to reach a state of zero volume and of infinitely great density. For example, when $k = 0$,

$$S = (1 - t/T)^{2/3},$$

and the time of collapse is

$$T = (2/3)(8\pi G\rho_0/3)^{-1/2}.$$

Similar values of T are found when $k = +1$ or $k = -1$; they are always very short for any astronomically probable values of ρ_0.

Binding energy. The radial coordinate r used in field (1) is comoving. Therefore the coordinate-radius r_b of the outer boundary

is fixed throughout the motion. Moreover zero-pressure collapse is characterized by the condition that

$$\rho S^3 = \text{constant} = \rho_0.$$

The mass of the configuration calculated from the internal field (1) is therefore constant and is

$$M = 4\pi\rho_0 \int_0^{r_b} \left(1 - k\frac{r^2}{R_0^2}\right)^{-1/2} r^2 dr. \tag{11}$$

On the other hand, it is possible to fit the external field (2) onto the internal (1) by finding the transformation that converts $(\bar{t}, u)$ into coordinates (t, r) in terms of which the metric (2) has a form similar to (1). This process leads to the conclusion that, for all three values of k,

$$m = (4/3)\,\pi G\rho_0 r_b^3/c^2. \tag{12}$$

The quantity

$$E = GM/c^2 - m$$

has been called the *binding energy* of the configuration by Fowler. It has the following expressions:

(i) If $\alpha - \beta^2 > 0$, or $k = +1$, then

$$E = \left(\frac{2\pi G\rho_0}{c^2}\right) \frac{c^3T^3}{(\alpha-\beta^2)^{3/2}} \left\{\omega_b - \sin\omega_b\cos\omega_b - \frac{2}{3}\sin^3\omega_b\right\}, \tag{13}$$

$$r_b = \{cT/(\alpha-\beta^2)^{1/2}\}\sin\omega_b.$$

(ii) If $\alpha - \beta^2 = 0$, or $k = 0$, then

$$E = 0. \tag{14}$$

(iii) If $\alpha - \beta^2 < 0$, or $k = -1$, then

$$E = \left(\frac{2\pi G\rho_0}{c^2}\right) \frac{c^3T^3}{(\beta^2-\alpha)^{3/2}} \left\{\sinh\omega_b\cosh\omega_b - \omega_b - \frac{2}{3}\sinh^3\omega_b\right\}, \tag{15}$$

$$r_b = \{cT/(\beta^2-\alpha)^{1/2}\}\sinh\omega_b.$$

In Case (i) the binding energy is always positive and continuously increases from zero with ω_b; in Case (iii) it is always negative and decreases continuously from zero as ω_b increases. In Case (ii), there is no binding energy. These motions of collapse therefore possess a positive, zero, or negative binding energy according to the

relative magnitudes of the initial density, proportional to α, and of the square of the initial rate of change of the scale-factor S.

Collapse with pressure-gradient. The zero-pressure condition seems to be very artificial from the physical point of view. However, there does exist an exact solution of Einstein's equations in which the pressure is zero at the boundary but increases inwards (McVittie [2]). It is derived from the following assumptions: (a) the configuration possesses spherical symmetry; (b) co-moving coordinates are possible; (c) there is a Schwarzschild type singularity at the center; (d) there is material outside the singular region whose stress-energy tensor reduces to its diagonal components, those which represent the stress being equal to one another. The metric of the internal field is

$$ds^2 = \left(\frac{1 - Q/S}{1 + Q/S}\right)^2 dt^2 - \frac{(1 + Q/S)^4}{c^2} R_0^2 S^2 \{ d\omega^2 + \mathscr{S}_k^2(\omega)\, d\Omega^2 \}, \tag{16}$$

where R_0 is a constant, S is a function of t, the function $\mathscr{S}_k(\omega)$ is $\sin\omega$, ω, or $\sinh\omega$ according as $k = +1, 0$ or -1, and

$$Q(\omega) = \frac{1}{4}\mu_0 / \mathscr{S}_k\left(\frac{\omega}{2}\right), \tag{17}$$

where μ_0 is a constant. If $\mu_0 = 0$, it is easy to show that (16) can be transformed into (1) by a transformation of the radial co-ordinate. If S is a constant equal to unity, $k = 0$ and $m = \mu_0 R_0$, $r = R_0\omega$ then (16) reduces to the isotropic form of the Schwarzschild field (2). Thus (16) gives the field inside a distribution of matter which is collapsing onto a central particle represented by a Schwarzschild singularity centered on the point round $\omega = 0$. The density and pressure are given by

$$\frac{8\pi G}{3}\rho = k\frac{c^2}{R_0^2}\frac{S^3}{(S+Q)^5} + \left(\frac{S'}{S}\right)^2, \tag{18}$$

$$8\pi G\frac{p}{c^2} = \frac{S}{S-Q}\left\{ -2\frac{(S+Q)}{S}\frac{S''}{S} - \frac{S-5Q}{S}\left(\frac{S'}{S}\right)^2 - k\frac{c^2}{R_0^2}\frac{S^3}{(S+Q)^5}\right\}, \tag{19}$$

where a prime now means a derivative with respect to t. Since the ω-coordinate is co-moving, its boundary value ω_b is fixed through-

out the collapse. Hence the condition $p = 0$ at the boundary is satisfied if the expression in curly brackets on the right-hand side of (19) is set equal to zero and Q_b is written for Q in the resulting equation. This equation then determines the function S. The properties of the motion are described elsewhere [2]; they do have some similarities with the $p = 0$ cases. But the attainment of an infinitesimally small volume now occurs, theoretically at least, only at the end of an infinite time. If it is assumed that the collapsing material obeys the perfect gas law $p = R\rho\theta$, where θ is the temperature, then it can be shown that the configuration rapidly becomes a very hot one. Temperatures of the order of 10^{10}° K are soon attained near the physical boundary of the central particle.

References

1. F. Hoyle and W. A. Fowler, Nature **197** (1963), 153.
2. G. C. McVittie, Astrophys. J. **143** (1966), 682.

UNIVERSITY OF ILLINOIS OBSERVATORY
URBANA, ILLINOIS

Jeffrey M. Cohen *

Dragging of Inertial Frames by Rotating Masses

In 1918, H. Thirring [1] showed that a slowly rotating mass shell (producing a weak gravitational field) has a weak dragging effect on the inertial frames within it. Thirring obtained his result by using the weak field equations obtained by Einstein [2] in 1916. These equations are obtained by assuming

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \tag{1}$$

where $\eta_{\mu\nu}$ is the Lorentz metric and $h_{\mu\nu}$ is a first-order quantity. By defining

$$\phi_{\mu}{}^{\nu} = h_{\mu}{}^{\nu} - \delta_{\mu}{}^{\nu} h/2 \tag{2}$$

with the coordinate condition

$$\phi_{\mu}{}^{\nu}{}_{;\nu} = 0, \tag{3}$$

one obtains the weak field equations

$$\Box \phi_{\mu}{}^{\nu} = -16\pi T_{\mu}{}^{\nu}. \tag{4}$$

Here $\Box$ is the D'Alembertian operator. This Equation (4), solved via Green's functions, yields the metric valid to first order in ω:

* This paper is based on the work of D. Brill and J. Cohen. Supported in part by the National Science Foundation and the National Aeronautics and Space Administration.

$$
\begin{aligned}
ds^2 = &-(1-2m/r_0)\,dt^2 \\
&+(1+2m/r_0)(dr^2+r^2d\Theta^2+r^2\sin\Theta\,d\phi^2) \\
&-(8m\omega/3r_0)\,r^2\sin^2\Theta\,d\phi\,dt,
\end{aligned} \tag{5}
$$

ω being the shell angular velocity, m the shell mass, and r_0 the shell radius. Thus the angular velocity Ω of the inertial frames within the shell is given by

$$\Omega/\omega = 4m/3r_0. \tag{6}$$

Since 1918, various authors have stressed the importance of obtaining a strong field solution in order to see how Mach's principle enters into general relativity. If Mach's principle is contained in general relativity, there should be a limiting mass for which Ω approaches ω.

In analogy with the isotropic Schwarzschild metric and Thirring's famous solution (5), consider the metric

$$ds^2 = -V^2dt^2+\psi^4(dr^2+r^2d\Theta^2+r^2\sin^2\Theta(d\phi-\Omega(r)\,dt)^2). \tag{7}$$

Here we have $\psi = 1+\alpha/r$ and $V=(r-\alpha)/(r+\alpha)$ for $r>r_0$ and $\psi = 1+\alpha/r_0$ and $V=(r_0-\alpha)/(r_0+\alpha)$ for $r<r_0$, where $\alpha = m/2$ = the Schwarzschild radius. For a slowly rotating mass shell producing a strong gravitational field, the nonvanishing components of the stress-energy tensor in orthonormal Cartan frames are

$$T^{00}, T^{22}, T^{33}, T^{03}$$

to first order in ω.

It is convenient to formulate this problem as a Cauchy problem. To first order in ω, the initial value equation

$$G^{00} = 8\pi T^{00} \tag{8}$$

is the same as that for the Schwarzschild metric. This equation gives the relation between the mass seen by an observer at infinity and that seen by an observer at the shell.

The only remaining nontrivial initial value equation is

$$G^{03} = 8\pi T^{03}. \tag{9}$$

The function Ω is determined by integrating this initial value Equation (9) across the thin shell, and requiring (a) regularity at the origin, (b) continuity of g_{03} across the shell, and (c) that the

metric reduce to the Schwarzschild metric for $\omega = 0$. To complete the Cauchy problem we note from the time development equations that the metric is time independent if $T^{22} = T^{33} = \beta T^{00}$ with $\beta = \alpha/2(r_0 - \alpha)$. Thus the initial value solution for Ω, valid for all time, is

$$\Omega = \frac{\omega}{1 + [3(r_0 - \alpha)/4m(1 + \beta)]} \tag{10}$$

for $r < r_0$. Here β results from elastic stresses in the shell.

For large r_0, this result (10) reduces to that of Thirring. On the other hand, as the shell mass increases, the rotating shell has an increasing dragging effect upon the inertial frames within it. Finally as $m \to 2r_0$, the shell rotation completely determines that of the inertial frames within it (i.e., $\Omega \to \omega$) while the effect of the inertial frames at infinity upon those within the shell goes to zero.

Comparison of the above solution with that of Kerr [3] shows that for small a and arbitrary m, Kerr's exterior solution can be matched to an interior solution.

References

1. H. Thirring, Phys. ZS. 19 (1918), 33.
2. A. Einstein, S.-B. Preuss. Akad. Wiss. (32) 69 (1916), 688.
3. R. Kerr, Phys. Rev. Lett. 11 (1963), 237.
4. R. Boyer, Proc. Cambridge Philos. Soc. 61 (1965), 527.

YALE UNIVERSITY
NEW HAVEN, CONNECTICUT

II. COSMOLOGY

Maarten Schmidt

Lectures on Quasi-Stellar Objects

I. The first lecture is to be about the observations of the quasi-stellar objects, and the second about their nature. In this lecture it will be assumed that they are very distant extragalactic objects, although in the second lecture some recent reasons for doubting this will be mentioned.

Of the radio sources in the Third Cambridge (3C) Catalogue, probably 50% can be identified with normal radio galaxies, whose optical images have a characteristic fuzzy appearance. Perhaps a further 25% will be identified with star-like objects, known as quasi-stellar radio sources (QSRS).

The first slide[1] illustrates the spectra of normal radio galaxies (3C445, 3C327, 3C234, 3C171). They show continuous emission with a strong superposed pattern of emission lines, Hβ and λ3727 of [O II] being conspicuous, while the wavelengths of the lines in

[1] Most of the slides were either of spectra (which could not be reproduced satisfactorily) or of material which was outdated at the time of this publication. Therefore there has been no attempt to reproduce the slides here.

the pattern depend on the redshift of the galaxy. Redshift is defined by

$$z = (\lambda - \lambda_{\text{rest}}) / \lambda_{\text{rest}}$$

where λ_{rest} is the rest wavelength of a line observed at wavelength λ. The decreasing intensity of the continuum with increasing redshift illustrates the redshift-magnitude relation which is found for radio galaxies. These emission spectra are similar to those of H II regions such as planetary nebulae, which are fully understood qualitatively, and reasonably well quantitatively. The appearance of forbidden lines indicates a low gas density, and the relative intensities of the lines give information about the distribution of ionization and excitation levels in the gas.

Slide 2 [1] shows the spectrum of the quasi-stellar radio source 3C273, from λ3500 to λ6300, widened to show up the low contrast emission features. Broad emission lines are visible with difficulty, and so is a narrow night sky line. From the ratios of their wavelengths, four of the lines can be identified as Hβ, Hγ, Hδ, Hϵ of the Balmer series with a redshift of 16%. This redshift leads to identification of a fifth line as λ5007 of [O III], which is also found in normal radio galaxies although with a different intensity relative to the hydrogen lines. This confirms the 16% redshift interpretation with absolute certainty.

3C273 has the optical appearance of a thirteenth magnitude star. However, the 16% redshift corresponds to a velocity greater than the velocity of escape from the Galaxy, so it would be unique in this respect if it were a star. In the relation between redshift and magnitude for normal radio galaxies, which is interpreted as a velocity-distance relation showing the expansion of the universe, a redshift 0.16 corresponds on the average to a 17th magnitude galaxy. Therefore if the redshift of 3C273 is a cosmological redshift, this source is four magnitudes (40 times) brighter than the most luminous normal galaxies. Also 3C273 is much smaller than a normal galaxy: the stellar image indicates a diameter less than $0''.5$, whereas a normal 17^{m} galaxy has a diameter $10''$ to $15''$.

Objects such as 3C273 are called quasi-stellar radio sources (QSRS).

Finding of QSRS has been made possible by radio observations of position to an accuracy of about $10''$. Occultation work enables

radio positions to be determined to about 1″, and it was this technique which enabled a definitive optical identification of the source 3C273 to be made [2], [3].

Slide 3 illustrates four QSRS: 3C273 [1] is a stellar object with a jet 1″ to 2″ wide at 11″ to 20″ from the star, which is a unique feature. The radio source is double: component B is 1″ west of center of the stellar object and of diameter less than 0″.5, while component A is 1″ west of the end of the jet and is elliptical, extending in the direction of the jet. This is unique among double radio sources in that the two components have very different spectra, and in that the components coincide with two optical features, since if a normal radio galaxy is a double source, the optical galaxy lies between the components.

3C48 is a stellar object surrounded by wisps of nebulosity. 3C196 and 3C147 have almost completely stellar images, with diameters less than 1″.

The optical properties of the QSRS are listed on slide 4.

1. They are starlike objects identified with radio sources. This distinguishes the QSRS from other recently discovered distant starlike objects [4], which are probably more numerous than the QSRS but are not radio sources.

2. Variability in light has been established for each of the five objects that have yet been carefully observed. Variations of about 30% in one or two years are found, and 3C273 shows indications of semiperiodic variation [5], [6]. Such variability is unexpected in extragalactic objects since extended galaxies are too large for large fractional variation to be possible (except when a supernova outburst occurs, which may give as much light as the whole galaxy for a brief time).

3. They have a large ultraviolet excess. This has been of practical significance for finding the QSRS. Sandage has identified many before accurate radio positions were known, by taking two exposures, one in ultraviolet and one in blue light, of the region, displacing the telescope slightly between exposures. The QSRS have much greater ultraviolet intensity relative to blue than normal stars of the same apparent magnitude. To date, about 70 QSRS have been identified, mostly by Sandage and Veron (at Palomar) and Wyndham (at California Institute of Technology). These are almost all in the Northern hemisphere.

4. Broad emission lines appear on most of the spectra observed. Of Schmidt's 23 spectra of QSRS, 3 show no emission lines, 7 show one line, and 13 show two or more. Of these 13, 9 have published values of the redshift [7] and the other 4 have unpublished redshifts in the same range. The emission lines are very broad, about 40A° total half-width, which corresponds to a velocity spread of about 3000 km/sec.

5. They have large redshifts. This property is of fundamental importance, as the QSRS have much larger redshifts than average normal galaxies of the same apparent brightness. The redshifts are very difficult to measure because of the low contrast between emission lines and continuum, quite unlike the spectra of normal radio galaxies.

Slide 5 shows four successive spectra of 3C273, unwidened. This is a 13^m object but the emission lines are difficult to see, even on the original plates, with an unwidened spectrum. For objects fainter than 3C273 widening of the spectrum is not possible because exposure times are limited by the night sky: even with good seeing, a four hour exposure must be made at 400A°/mm, and longer exposures are necessary at poor seeing. For faint objects, the existence of lines has been checked by taking repeated unwidened exposures.

Slide 6 shows the spectrum of 3C48, which is a $16^m.5$ object with a redshift 0.37. Numerous lines have been identified, and two can be seen on the slide: λ3727, a forbidden line of singly ionized oxygen, and λ2798 of ionized magnesium. The spectrum is discussed by Greenstein and Schmidt [1].

Slide 7 shows the spectrum of 3C147, with the lines emphasized by using a composite of six negatives; the redshift is $z = 0.54$ [8].

Slide 8 shows the spectrum of 3C254. Five lines have been found, using four or five exposures. If the redshift is taken as 0.734, four of the lines occur exactly at the redshifted positions of λ3869 of [Ne III], λ3727 of [O II], λ3426 of [Ne V], and λ2798 of [Mg II]. Accepting this as the correct identification, one finds the fifth line corresponds to λ1909 of C III, thus establishing the presence of this emission in a QSRS. The spectra of two more sources, each showing only two lines with wavelength ratios 1.468 and 1.467, can then be interpreted if these lines are identified as λ2798 and λ1909. The corresponding redshifts are about unity for each source—i.e.,

the observed wavelengths are about twice the rest wavelengths.

Slide 9 shows the spectrum of 3C287 from λ3188 to λ5016. Three lines are visible. Two have wavelength ratio 1.467, and if these are taken to be λ2798 and λ1909 the third line corresponds to λ1550, an emission of C IV.

Slide 10 shows the spectrum of 3C9 [7] in the blue region. Only two lines have yet been found, both very broad, and the wavelength ratio, 1.27, does not correspond to any pair of lines yet identified in other sources. On the basis of the theory of planetary nebulae, a list has been prepared of emission lines that may be expected. Inspection of this list shows that the only reasonable pair of lines with wavelength ratio 1.27 is λ1550 of C IV and the Lyman α line λ1216. This is the proposed identification of the lines in the spectrum of 3C9. It leads to a redshift of 2.01: the lines are observed at a wavelength more than three times the rest wavelength.

The spectrum of 3C9 has a further very significant feature: continuum emission at wavelengths below Lyman α is visible, down to a rest wavelength about λ1100. As will be discussed further in the second lecture, this observation sets an extremely low upper limit to the density of neutral atomic hydrogen in the intergalactic medium. The continuum intensity is actually observed to drop rapidly with decreasing wavelength below the Lyman α line, but this effect may arise in the earth's atmosphere so care is needed in its interpretation.

Slide 11 shows the optical appearance of 3C9. It is a very blue 18^m star which is listed in the Luyten-Haro Catalogue of faint blue stars. Also shown is Sandage's object BSO1, which has the appearance of a faint blue star and a redshift probably similar to 3C9, but is not observed as a radio source [4].

If the redshifts of the QSRS are cosmological, one can in principle from the redshift-magnitude relation obtain information about the geometry of the universe. There is so much uncertainty as to the nature of the objects that it may not in practice be possible to draw significant conclusions for cosmology; but ten years ago cosmology was being studied with observations out to redshifts of 0.2, so certainly there is a better chance now that redshifts of 2 are observable.

The QSRS are of great interest in themselves, whether they can be used effectively for cosmology or not. They may, for example,

be related to the radio galaxies and have related energy sources. It will be assumed here that they are at the very great extragalactic distances indicated by cosmological redshifts, and a few of the very large number of proposed models will be discussed.

The presence of the Balmer series, identified in the nearer objects, and forbidden lines shows that H II regions (ionized gas) are present. An estimate of the total mass of ionized gas can be made in the following way. (Details are given in the paper by Greenstein and Schmidt [1].) Computations of the energy output from an H II region show that the total energy emitted in the Hβ line is approximately

$$E(\mathrm{H}\beta) \doteq 10^{-25} N_e^2 \,\mathrm{ergs\,sec^{-1}cm^{-3}},$$

where N_e = number of electrons per cm^3. (The numerical value depends, not sensitively, on an assumed electron temperature.) For 3C273, direct observation gives the flux density $F(\mathrm{H}\beta)$ in ergs sec^{-1}cm^{-2}, and the redshift gives its distance, r. Thus the luminosity in Hβ, in ergs sec^{-1}, can be calculated, and related to the volume of gas, V cm^3, by

$$F(\mathrm{H}\beta) \cdot 4\pi r^2 = 10^{-25} N_e^2 V.$$

An estimate of N_e can be obtained by observations of the forbidden lines. For a forbidden line to be formed, the upper level is populated by collisional excitation, and deexcitation must occur by spontaneous emission; if the density is too great, collisional deexcitation will dominate. (More precisely, if N_e is large enough the intensity of a forbidden line increases as N_e while that of a permitted line increases as N_e^2.) In 3C273, the absence of λ3727 of [O II] shows that $N_e > 3 \times 10^4 \mathrm{cm}^{-3}$. Similarly, the presence of a weak line at λ5007 of [O III] shows that $N_e < 10^7$. Adopting the value $N_e \doteq 3 \times 10^6 \mathrm{cm}^{-3}$ gives for the volume of ionized gas emitting the lines, $V = 2 \times 10^{55} \mathrm{cm}^3$. If this is a sphere, its radius has the very small value of about 1 parsec.

An upper limit to the size of the region producing the optical continuum may be obtained from observations of the light variations, which seem to have a periodic component about ten years [5], [6]. Some authors regard the variations as due to random outbursts (e.g. uncorrelated super-supernova explosions), in which

case any size is possible for the object. Greenstein and Schmidt [1] follow Smith's interpretation, setting an upper limit to the radius at one light-year. This is only one third of the radius of the H II region, if the latter is a sphere of 2×10^{55}cm^3 volume. But because the optical depth for scattering through such an H II region is large, each photon must be scattered many times as it travels out to 1 parsec radius, and the fluctuations of light at the surface cannot be much more rapid than if the continuum source itself were of 1 parsec radius.

A filamentary structure for the H II region could lead to the required optical properties: within the filaments N_e would be between 10^6 and 10^7cm^{-3}, the volume of the filaments would be 2×10^{55}cm^3, but these would be spread over a much larger total volume and so have low opacity. In this case rapid fluctuations of a small central source could be observed externally.

Until recently it has been possible to explain the radio observations in terms of the filamentary model as follows. If the radio emission arose within the region occupied by the filaments, Faraday rotation would destroy all polarization and the region would be optically thick for free-free radio emission. Both these effects are contrary to observation. It may be supposed instead that the radio emission comes from a larger surrounding medium, which would also be in agreement with the occultation data that set an upper limit of 0″.5, or 500pc, to the size of the radio source. However, Dent [9] has recently discovered a totally unexpected variability in the radio output of 3C273: the flux at $3\frac{1}{2}$ cm wavelength has increased 40% in three years. This makes a size larger than a few light-years improbable for the radio source, so the model discussed above may have to be rejected.

II. Table I [7] lists the QSRS with known redshifts and gives in the third column the calculated intrinsic power emitted by the sources at λ3000, in watts (cycle/sec)$^{-1}$. This is calculated on the assumption of an exploding cosmological model, with the parameters: Hubble's constant $H_0 = \dot{R}_0/R_0 = 100\,\mathrm{km\,sec^{-1}Mpc^{-1}}$, deceleration parameter $q_0 = -\ddot{R}_0 R_0/\dot{R}_0^2 = +1$, cosmological constant $\Lambda = 0$, corresponding to a mean density $4 \times 10^{-29}\,\mathrm{gm\,cm^{-3}}$. The luminosities range over a factor 20, with a mean deviation corresponding to $0^m.85$.

TABLE I

Redshifts and Intrinsic Luminosities of QSRS

Source	Redshift (z)	$\log F(10^{15}\mathrm{c/s})$, F in watts $(\mathrm{c/s})^{-1}$ at λ3000, for cosmological model
3C273	0.158	23.8
3C48	0.367	23.0
3C47	0.425	22.5
3C147	0.545	23.1
3C254	0.734	22.9
3C245	1.029	23.5
CTA102	1.037	23.5
3C287	1.055	23.4
3C9	2.012	23.6

The significance will now be discussed of the continuum radiation seen in the spectrum of 3C9, corresponding to rest wavelengths down to λ1100, below Lyman α. The radiation emitted by 3C9 at λ1100 passes on its way to the earth a point in the universe at which it appears redshifted to λ1216, at which point resonant scattering in Lyman α by neutral hydrogen would occur. From the observed intensity of the light reaching the earth, Gunn and Peterson [10] have deduced that an upper limit for the density of neutral atomic hydrogen at that point in the universe is 3×10^{-11} atoms per cm^3. That point is at a distance from the earth corresponding to a redshift of about 2, and a time when the cosmic density was about 27 times greater than the present density. Therefore the corresponding upper limit to the density of neutral atomic hydrogen in the present intergalactic medium is about $10^{-12}\mathrm{cm}^{-3}$. An important consequence of this extremely low value is that essentially all the intergalactic hydrogen must be ionized, so electron scattering must be an important source of opacity. For example, for a cosmological model with $q_0 = +1$, the optical depth in electron scattering between 3C9 and earth is about 0.5. This effect must be considered in future interpretations of the redshift-magnitude relation.

For the rest of the lecture, theories as to the nature of the redshifts of the QSRS will be discussed.

The first possibility considered was that the redshifts of 3C273 and 3C48 were gravitational [1]. For a star of mass M and radius R, the gravitational redshift is

$$z = \Delta\lambda/\lambda \cong GM/Rc^2.$$

E.g., 3C273 could be a neutron star of solar mass and about 10 km radius. The depth of the atmospheric shell emitting the lines can be calculated by supposing that the line widths correspond to the change in gravitational potential across the shell: for lines of width w the shell thickness ΔR is given by $\Delta R/R = w/\Delta\lambda$. The line widths are 7% of $\Delta\lambda$ in 3C273 and 1.6% in 3C48. Thus if the redshift of 3C48 is gravitational, its distance can be estimated for any assumed mass, as follows.

Assuming a value for the mass, the radius is determined from the observed value of z, and then the shell thickness ΔR is determined from the observed line width w, which gives $\Delta R = 0.016R$. Then the volume V of emitting gas is $4\pi R^2 \Delta R$. The presence of λ3727 of [O II] at normal strength relative to the Balmer lines shows that N_e is no greater than $3 \times 10^4 \mathrm{cm}^{-3}$. With this value of N_e, a value for the luminosity in Hβ is given by $L(\mathrm{H}\beta) = 10^{-25} N_e^2 V \,\mathrm{ergs\,sec}^{-1}$. But the observed flux in Hβ is related to L and the distance r, since $L(\mathrm{H}\beta) = 4\pi r^2 F(\mathrm{H}\beta)$. In this way, a value of r is calculated for each assumed mass.

This has been done for the range one to 10^{14} solar masses. If the mass is $1M_\odot$, r is found to be 14 km! An object of mass $10^4 M_\odot$ would still be within the solar system, at 0.1 A.U., so is clearly impossible. An object of $10^{10} M_\odot$ is also impossible, since it would be at 0.5 kpc and have a gravitational effect 36 times that of the whole galaxy. If masses as large as $10^{14} M_\odot$ are considered, the distances are so great that the redshifts would be cosmological.

In fact the situation is even less favorable, because taking $\Delta R/R = w/\Delta\lambda = 0.016$ was an overestimate. Radiation effectively originates from an atmospheric layer not thicker than about one scale height; an upper limit to the scale height can be set because the temperature cannot be greater than that corresponding to the line widths, and this shows that $\Delta R/R = 10^{-5}$ should be the value used. Revised calculations with this shell thickness show that a mass of $10^{13} M_\odot$

would have to be at 0.3 Mpc. This is ruled out by observed motions in the Local Group of galaxies, which show its total mass to be about $2 \times 10^{12} M_{\odot}$.

Summarizing, if the redshifts are gravitational in origin, 3C48 is not a neutron star and it is not a galactic object. It must be more distant than 2 Mpc and more massive than $2 \times 10^{13} M_{\odot}$, which is very unlikely.

A theoretical argument against the gravitational redshift hypothesis is due to Bondi [11]: a collapsed static object with plausible properties cannot have a gravitational redshift greater than 1.3, which is smaller than the largest redshift observed. A further argument is the existence of a redshift-magnitude relation for the objects. Although the relation cannot be precisely specified from only 9 redshifts (especially since there is observational selection of the intrinsically brightest objects at large redshifts), the data in Table I do indicate a normal redshift-magnitude relation and the newest results are in agreement. Such a relation is not explained if the redshifts are gravitational.

A second possible explanation of the redshifts is that they represent local Doppler shifts. Difficulties are the very high velocities (up to 0.8c has been observed) and the absence of any observed blueshifts. Of the 23 sources whose spectra have been examined, 13 show definite redshifts, 7 more have one spectral line which can be interpreted as redshifted $\lambda 2800$, and only three show no spectral lines. Therefore only a small fraction of the sources could have velocities of approach, and an explosive event with a local center is suggested.

The solar system is not a possible center, but it has been suggested by Terrell [12] that the objects were shot out by an explosion at the center of the Galaxy. Using this hypothesis, it is possible to estimate the masses of the objects. Last year it was thought that $1000 M_{\odot}$ would suffice, but a revised estimate, as follows, gives about $10^6 M_{\odot}$ for 3C273.

Two cases are considered, a stable or unstable object. If the object is stable, its mass is approximately $M \sim G^{-1} R (\Delta v)^2$, where G is the gravitational constant, R is the radius, and Δv is the velocity dispersion indicated by the widths of the emission lines. For 3C273, $\Delta v \doteq 1000$ km/sec, and the model of Greenstein and Schmidt [1]

discussed in the first lecture gives a value of R depending on the distance, r: $4\pi r^2 F(\mathrm{H}\beta) = 10^{-25} N_e^2 \cdot 4\pi R^3/3$. The mass is thus given by $M = 3 \times 10^6 M_\odot (r/1\,\mathrm{Mpc})^{2/3}$.

If the object is unstable, it is losing mass by expansion at a rate given by

$$\text{mass loss} = \Delta v \cdot 4\pi R^2 N_e \cdot 2 \times 10^{-24} \text{gm sec}^{-1}.$$

The time involved, if the speed has been constant, is $T = r/v$, and as before R is expressed in terms of r. In this case the mass is estimated to be about equal to the mass lost in time T, which is $M = 5 \times 10^6 M_\odot (r/1\,\mathrm{Mpc})^{7/3}$.

A lower limit to the distance is set by the very accurate proper motion studies by Jeffreys [13]. He finds $\mu = 0.001 \pm .0025''$/year. Combining this with the known radial velocity relative to the sun (given by z) and assuming that 3C273 is moving radially from the Galactic center, gives a minimum distance of 0.2 Mpc. Therefore it appears that the mass of 3C273 cannot be less than about $10^6 M_\odot$.

The total energy involved in the Galactic explosion can now be estimated. Probably 1000 or more QSRS exist, since 70 have been identified already, all in the northern hemisphere. This gives a total mass about $10^9 M_\odot$. Nearly all the objects are very faint and must have redshifts similar to 3C9, so their total kinetic energy must be at least of the order of $10^9 M_\odot c^2 \doteq 10^{63}$ergs. The time of the explosion, from the distance and velocity of 3C273, is about 20 million years ago. Even with 1% efficiency, there must have been, only 20 million years ago, an explosion at the Galactic center involving energy equal to the rest mass energy of the entire Galaxy, which shot out 1000 objects with a total kinetic energy of 10^{63}ergs! Even more energy is required if the quasi-stellar galaxies were also ejected, since there are 500 times as many of these per unit volume as the QSRS [4].

The occurrence of such an explosion, so recently, at the center of a quiescent galaxy like ours is extremely improbable; the energy requirements provide a very strong argument against the hypothesis of local Doppler redshifts.

Hoyle and Burbidge [14] have a new theory for the QSRS involving local Doppler shifts. They consider that the explosion could have occurred in a nearby radio galaxy, which is a preferable

hypothesis since energies of 10^{63} to 10^{64}ergs are already involved in the normal radio output. An explosion in the peculiar radio galaxy NGC5128 is suggested. It is only 4 Mpc away, and the expected number of approaching objects is not so great that blue-shifts should have been observed. NGC5128 is 45° away from 3C273, so from the observed proper motion of 3C273, supposing it moves radially away from NGC5128, it is concluded that its distance is 8 to 10 Mpc. Therefore the mass of 3C273 is at least five times greater than in the case of a local Galactic explosion. If the quasi-stellar galaxies arose in the same way, the total energy requirement is about 10^{65}ergs, which is the rest mass equivalent of a galaxy!

Finally, there is the possibility that the redshifts observed are cosmological. Four objections to this hypothesis will be considered.

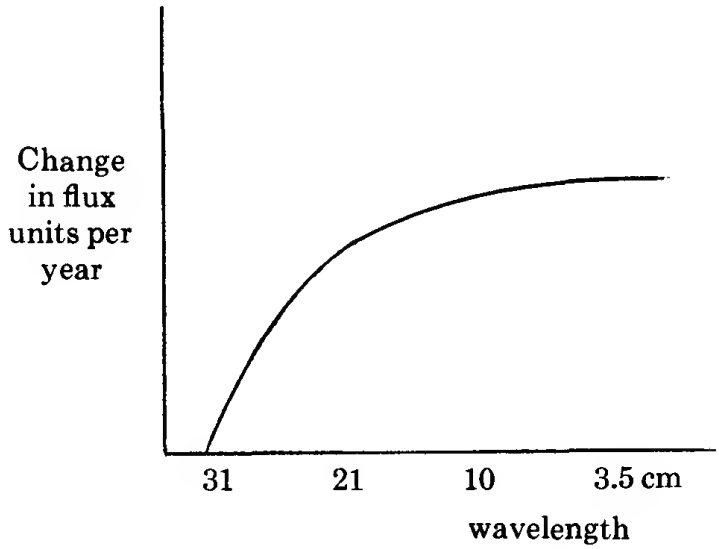
The first objection is that QSRS are not seen in clusters of galaxies. This is not a strong argument because there are so few observations. Since the QSRS are about four magnitudes brighter than normal galaxies, one must look to four magnitudes fainter to observe cluster members. This has been done for 3C273 and 3C48, and no members found, and although no others have been closely examined, it is felt that cluster members would have shown up. Among normal radio galaxies one third to one half occur in clusters, so this argument shows only that the QSRS may not be similar to the normal radio galaxies.

A second objection is that QSRS are not at the center of normal galaxies. This is not really a strong objection either, although one would have been happier if the sources had been found to be bright galactic nuclei. Probably it is a photographic impossibility to bring out a galaxy 5″ to 10″ round a stellar object that is 30 to 100 times brighter.

A third difficulty with cosmological distances is the variability in light. This has been discussed in the first lecture, and it is not a very great difficulty since the object as a whole need not be of very small size.

The fourth objection is much more serious: the radio variability. Dent [9] recently discovered that at 3.5 cm wavelength the flux from 3C273 increased by 40% in 1000 days. In reporting on this observation, Dent gave two strong arguments against the cosmological distance for 3C273. If the radio variation indicates a

change over the whole of a homogeneous body, the diameter of the radio source must be no more than 2 light-years. Using Hazard's [2] angular diameter of 0″.5 gives a distance only 10^6 light-years, which is only 1/200 of the cosmological distance and is within the Local Group. This angular diameter may not be relevant as it was measured at a different frequency, but Dent's second argument concerns the radio spectrum. If 3C273 were a source of 2 light-years diameter, radiating by the synchrotron mechanism at a rate given by assuming the cosmological distance, the radio spectrum would vary with frequency as $\nu^{2.5}$. However, the observed spectrum of 3C273B is flat.



There are several reasons why Dent's observations need not be regarded as conclusively showing that 3C273 is closer than the cosmological distance. Hazard has suggested that 0″.5 may not be the correct angular diameter as the occultation data are very difficult to interpret. Hoyle and Burbidge [14] show that if an unusual and intense magnetic field configuration is present, in which the spectrum is determined by the distribution of energy in the magnetic field instead of in the particles, then a very small size is just theoretically possible. Finally, Maltby and Moffet [15] have made a significant determination of the spectrum of the radio variation of 3C273B, from a study of old records. They find that the variation is much smaller at 31 cm. Since 3C273B itself has

a flat radio spectrum, this proves that it is not a representative part of the source which varies by 40%. Possibly a small region nearby is variable at 3.5 cm; this region must be optically thick in the 25 cm range, which is consistent with a small size. Certainly Dent's observation does not prove that 3C273B itself has a diameter of a few light-years.

Points in favor of cosmological distances for the QSRS will now be mentioned.

First, other explanations are unlikely.

Second, there is the slope of the count-magnitude relation for the quasi-stellar galaxies [4]. The slope of a graph of $\log N(\mathrm{m})$ versus m for blue stars at high galactic latitudes increases greatly at 15^{m}, and for fainter magnitudes has the value $d \log N(m)/\mathrm{dm} = 0.383$. This indicates that the objects fainter than 15^{m} are extragalactic. There are two lines of supporting evidence—these stars are relatively brighter in the ultraviolet than those below 15^{m}, and of the three spectra obtained for the faint blue stars one shows a large redshift. Sandage shows that the slope 0.383 can be explained in terms of an exploding cosmology with $\Lambda = 0$ and q_0 between 0 and 1, if the objects are extragalactic with intrinsic luminosities similar to the QSRS. (In Euclidean space the slope would be 0.6.) Hoyle and Burbidge are able to explain this slope in terms of their explosion theory, by adopting a suitable luminosity and velocity distribution for the objects. In this case, their total mass is only about $10^7 M_\odot$, so the energy requirements for the explosion are not so prohibitive.

A third point in favor of cosmological distances for the QSRS is that they would have radio properties similar to the radio galaxies.

A fourth point is that Koehler [16] has radio observations suggesting 21 cm neutral hydrogen absorption in the spectrum of 3C273. The exact frequency of the absorption line corresponds to the velocity of the Virgo cluster, which is in the direction of 3C273, and suggests that the source is more distant than this cluster, i.e., 10 Mpc. However, this would still be consistent with the theory of Hoyle and Burbidge involving a nearby explosion, as well as with the suggestion of cosmological redshifts.

There are no doubt many other points which could be discussed concerning the nature of the quasi-stellar objects.

References

1. J. L. Greenstein and M. Schmidt, Astrophys. J. **140** (1964), 1.
2. C. Hazard, M. B. Mackey and A. J. Shimmins, Nature **197** (1963), 1037.
3. M. Schmidt, Nature **197** (1963), 1040.
4. A. R. Sandage, Astrophys. J. **141** (1965),
5. H. J. Smith and Dorrit Hoffleit, Nature **198** (1963), 650.
6. Robinson, Schild and Schücking, eds., *Quasi-stellar sources and gravitational collapse*, Univ. of Chicago Press, 1964; Chapter 16, H. J. Smith.
7. M. Schmidt, Astrophys. J. **141** (1965), 1295.
8. M. Schmidt and T. A. Matthews, Astrophys. J. **139** (1964), 781.
9. W. A. Dent, Science **148** (1965), 1485.
10. J. E. Gunn and B. A. Peterson, Astrophys. J. **142** (1965), 1633.
11. H. Bondi, Proc. Roy. Soc. Ser. A **282** (1964), 303.
12. J. Terrell, Science **145** (1964), 918.
13. W. H. Jeffreys, Astronom. J. **69** (1964), 255.
14. F. Hoyle and G. R. Burbidge, Astrophys. J. **144** (1966), 534.
15. P. Maltby and A. T. Moffet, Science **150** (1965), 63.
16. J. Koehler (unpublished).

California Institute of Technology
Pasadena, California

E. L. Schücking

Cosmology[1]

A. **Principles.** Cosmology is the science of the universe. As such it may be a metaphysical science for us since the universe might contain regions which are, in principle, unobservable for us for all times. Sometimes cosmology is defined somewhat more narrowly as the science of the universe that can be observed in principle. Only this science is, in principle, an empirical one.

The growth of cosmology and its scientification has paralleled the growth of the observable universe and its secularization. The Christian universe of the Holy Thomas was probably much smaller than 10^{16} cm in diameter and contained heaven and hell. The modern universe is larger than 10^{27} cm, and no scientist in his right mind believes nowadays that heaven or hell are regions of the observable four-dimensional space-time manifold. Although some cosmologists might still unconsciously project theological pictures into their science, we have been able to scare most of the ministers out of cosmology by a straightforward application of tensor analysis.

Among the many attitudes of the modern scientist towards cosmology, two deserve special mention, the local and the global point of view.

[1] This work was supported by the grant AF-AFOSR 454-64/65.

(a) *The local standpoint.* One assumes the local laws of physics as known and determines the structure of the universe from observation by extrapolation of the local laws. This is the same method we use when determining the structure of the earth, the solar system, or the structure of our galaxy. Since most of the laws of classical physics are given by hyperbolic differential equations, we use essentially observational data to fix the unknown functions in an initial value problem. The local laws describe then evolution of the system in time. This theory is not tailor-made for the universe.

(b) *The global standpoint.* One assumes a four-dimensional structure of the universe based on more or less appealing principles. One tries then to find local laws which give the known local ones approximately and allow the universe as an essentially unique solution of the new equations. Steady state, Jordan's, and Milne's cosmologies seem thus to be much more specific since they narrow down the infinite number of universes to essentially one. But since they make high symmetry assumptions that might only be fulfilled by the universe itself, testing of the theory on the local asymmetrical scene leads to difficulties.

Ideologically in this class belong also different principles invented by Dicke, Einstein, Hoyle, Sciama, Wheeler, and many others that have all been named after the unfortunate Ernst Mach, the godfather of Wolfgang Pauli, who brought so much clarity into physics and is now being credited with so much obscurity. Although we do not have a satisfactory theory based on the local or global point of view, I feel that the conventional local theories have to be pressed much harder before we could discard them. Theories of the global type will not be discussed in the following. We shall take the local point of view as is customary in astrophysics.

B. **Local laws.** We shall assume textbook physics and extrapolate it to universal validity. This includes, naturally, the assumptions that the constants of nature are true constants and not space-time-dependent scalar fields or, even more generally, components of space-time-dependent tensor fields. These assumptions may not be made arbitrarily, but they need to be checked. The constancy of the weak interaction (β-decay) is, for instance, involved if we obtain for a rock the same age with the potassium-argon method

as with the uranium-lead method. What one checks then precisely is that the ratio of a strong interaction to the weak interaction constant was unchanged in the neighborhood of the earth averaged over a certain geological time. This field of "experimental cosmology" is a very important foundation for all our speculations about the universe. Most laws of nature have limited validity only, and their universal application might lead to crucial errors. The infinite life-times of protons and electrons under ordinary conditions, the zero rest mass of the photon, the electrical neutrality of the hydrogen atom, the constancy of gravitational interaction, the proportionality of inertial and passive gravitational mass, or the scalar (not tensor) nature of mass have all been confirmed with sometimes staggering accuracy.

Assuming electric neutrality of the cosmic plasma (matter) and magnetic fields outside of galaxies weaker than inside, it would seem that gravitation is the overwhelming long range interaction which governs the large scale structure of matter and its motions. There is the possibility that there exist interactions in the cosmic realm based on zero rest mass fields, for instance, with spin 3, the source of which is instead of charge or energy, a property of matter we might call mysterity. Mysterity might have appreciable positive or negative values only for masses with more than 10^{40}g and high relative velocities or spins and, therefore, its measurement might not be a standard laboratory job. The interaction of two mysterities through a spin 3 rest mass zero field is only one of an infinite number of possibilities.[2] Observationally we know practically nothing about the question if, say, the interaction between two clusters of galaxies in a distance of 10 Mpc exists and how much gravitation contributes to it—if at all. Theoretically this idea has been discussed by assuming a finite rest mass for the graviton. This implies a range of the gravitational force of the order of the Compton wavelength h/mc of the graviton (m = hypothetical mass of hypothetical graviton).

We shall assume the validity of Einstein's theory of gravitation although only the spherically symmetric, static solution of this theory has been tested so far observationally. Imagine the analogous

[2] Compare, however, S. Weinberg, "The quantum theory of massless particles" in *Lectures on particles and field theory,* Prentice-Hall, N. J., 1965.

situation in Maxwell's theory where people should believe in Maxwell's equations before Volta, Oersted, Ampere, Faraday, Galvani had done their experiments, magnetism had not yet been observed in one instance, and there was only good old Coulomb with his $1/r^2$ law!

C. **Newton's theory.** Instead of Einstein's theory we shall first consider Newton's, which is a very good approximation to Einstein's under all circumstances of which we have positive knowledge. Newton's theory of gravitation has recently been rediscovered by a number of people, among whom I mention Kilmister, Havas, and Trautman. If you should have cherished the illusion that you know it, then especially a glance into Trautman's paper will immediately dispel it. He shows that in its invariant formulation on the local level the Newton theory is much more complicated than the Einstein theory. It is on the global level, however, where Newton's theory is simple and Einstein's theory practically not yet existent.

There are some good reasons to discuss Newton's theory here at some length: Its applications to cosmology are not so well known, and it is a good approximation to Einstein's theory.

1. *The Inertial System.* The student of physics is informed on the very first pages of almost every textbook that the laws of physics are valid in an inertial system. A satisfactory answer how such a system could be found is, however, never given by the authors. The walls of a laboratory on earth do not provide such a system due to the rotation of the earth and its gravitational field. A nonrotating system (no Coriolis forces) fixed in the center of the earth is subject to the gravitational attractions of moon and sun. A nonrotating system fixed in the center of mass of the solar system (Newton's solution to the problem, 1687) is accelerated by the field of the galaxy. A system fixed in the center of the galaxy is accelerated by the attraction of other galaxies, and so on.

This difficulty in the definition of an inertial system is caused by the presence of interacting matter in the universe whose interaction is proportional to the inertial mass. One could set up an inertial system by simply ignoring gravitation—as one does in laboratory physics. In cosmology, however, gravitation appears to be the predominant force. Switching it off or removing the

matter from the universe would make the cosmological scene completely obscure.

Another proposal dealing with this difficulty is the theory of the matter island universe. If the mass of the universe were finite, we could use the center of mass of the universe for the definition of an inertial system. This would not help us, however, for our operational definition of an inertial system. A thinning out of matter is not observed in the observable range, and our knowledge about the mass distribution in the universe is so scanty that such a definition would not be helpful for the determination of an inertial system even if the basic assumption of the island universe was correct.

It is, however, possible to uphold the notion of an inertial system even in the presence of gravitational fields in a universe that has an infinite mass. All one has to do is pick a particle that moves freely under the influence of inertia and gravitation. To this particle one attaches permanently a nonrotating system of coordinates. We call such a system a local inertial system. Describing a Newtonian potential by $\Phi = \Phi(x_j, t)$ we take as equation of motion for a particle under the influence of a nongravitational force K_i in the local inertial system

$$m\ddot{x}_i = K_i - m\phi_{,i}.$$

Since $x_i(t) = 0$ is a solution of this equation for $K_i = 0$ (particle in the origin), we have

$$\phi_{,i}(0, 0, 0, t) = 0.$$

Such a local inertial system can be tied to a space craft flying with its engine shut off and experiencing no air resistance.

A genuine gravitational field makes itself felt by *relative* accelerations of local inertial systems. One may think, for instance, of local inertial systems tied to particles falling radially towards the center of the earth from opposite directions. In the presence of a gravitational field, we can no longer transform from one local inertial system to another by means of transformations belonging to the Galilei group. The transformations that we shall employ going over from one local inertial system to another are found by requiring that the nongravitational force K_i $(i = 1, 2, 3)$ transforms like a vector. That implies constancy of the components for systems with parallel axes. We have for a particle

with coordinates x_i and x_i', respectively, in the two systems

$$K_i = m\left[\ddot{x}_i + \frac{\partial}{\partial x_i}\Phi(x_j, t)\right] = m\left[\ddot{x}_i' + \frac{\partial}{\partial x_i'}\Phi'(x_j', t)\right].$$

Calling $\ddot{x}_i'(t) - \ddot{x}_i(t) = g_i(t)$ the relative acceleration between the two local inertial systems we take

$$\Phi'(x_j', t) = \Phi\left[x_j(x_k'), t\right] - g_i(t)\, x_i(x_k') + h(t),$$

$h(t)$ can be taken to be $\Phi(0,t) + g_i(t)\, x_i(0) - \Phi\left[x_i(0), t\right]$. This insures that the gravitational potential has the same value (for instance, zero) in the origin of every local inertial system. This normalization is convenient especially in the cosmological case. The new feature in our formulation of Newton's theory is a transformation of the gravitational potential. This change in the potential amounts to adding the potential of a homogeneous gravitational field. The functions $g_i(t)$ which enter into the transformations depend now on the special choice of the gravitational potential Φ.

The transformation of the Newtonian potential leads to an interesting enlargement of the Galilei group. Let us assume that no true gravitational fields are present ($\rho = 0$ everywhere and $\phi_{,i,k} \to 0$ for $r \to \infty$). We consider the ∞^{16} different frames of reference which are related to each other by the transformations

$$x_i' = \alpha_{ik}(x_k + g_k t^2/2 + \beta_k t + \gamma_k), \qquad \alpha_{ik}\alpha_{ie} = \delta_{ke},$$

$$t' = t + \delta, \qquad \alpha_{ik}, \beta_k, \gamma_k, \delta, \epsilon, \eta, \lambda, g_i = \text{const.},$$

$$\Phi' = \Phi - g_i x_i + \epsilon t^2 + \eta t + \lambda.$$

This is a sixteen-dimensional Lie group operating as group of transformations on the four-dimensional space with coordinates x_k and t. This group contains the Galilei group as a subgroup. We shall call it the super-Galilei group. The sgp leaves the equations of motion

$$K_i' = m\left[\ddot{x}_i' + \frac{\partial \Phi'(x_j')}{\partial x_j'}\right] = m\left[\alpha_{ik}(\ddot{x}_k + g_k) + \frac{\partial x_k}{\partial x_i'}\left(\frac{\partial \Phi}{\partial x_k} - g_k\right)\right]$$

$$= m\,\alpha_{ik}\left[\ddot{x}_k + \frac{\partial \Phi}{\partial x_k}\right] = \alpha_{ik} K_k$$

invariant. The group allows us to make transformations from an inertial frame to a frame in which a homogeneous gravitational field is present and vice versa.

We consider now the gravitational potential as the fifth coordinate of an event. The group becomes a transformation group of a five-dimensional space with coordinates x_k, t and Φ. This is possible due to the universal coupling of the gravitational field to all matter. The introduction of the sgp with its transformation of the potential has some resemblance to the introduction of the Lorentz group that transforms time.

Two generalizations come immediately to one's mind:

(a) Lie groups that transform to rotating coordinate systems and other more general ones which give rise to more complicated nongenuine gravitational fields.

(b) Corresponding Lie groups which generalize the Poincaré group. What is the purpose of such investigations? We hope to find in this way coordinate systems that are described by a finite number of parameters instead of by four functions of four variables. Einstein's jump from the Poincaré group to the pseudogroup of all coordinate transformations would thus be interpolated by some intermediate steps. Such groups would introduce new quantities (gravitational potentials) which might also prove helpful for the description of genuine gravitational fields. It is also conceivable that such Lie groups could be useful for the introduction of global coordinate systems into curved space-time manifolds.

I do not know what a relativistic analogue of the sgp would look like. The Newtonian potential should in a relativistic version become the "oo"—component of a symmetric tensor. One might have to go then into a fourteen-dimensional space instead of a five-dimensional one. The combination of genuine Lorentz transformations and accelerations will give rise to Thomas precession. One has to include, therefore, rotating coordinate systems. The fifteen-dimensional conformal group does not seem to be the correct solution since its homothetic transformations do not appear in the sgp.

2. *The Friedman Models.* The Newtonian analogues of the Friedman models were given by Milne and McCrea in 1934. We call these models after Alexander Friedman, a Russian meteorologist and flying champion who was the first to propose an evolutionary world model in the frame of Einstein's theory of gravitation in 1922. Friedman's idea that the universe could evolve is one of the most important suggestions ever made in the history of science.

It is very easy to derive Friedman's differential equation for the evolution of the universe. We take two mass points in the universe at a distance $R(t)$. The acceleration of the second with respect to the first is then assumed to be

$$\ddot{R} = -GM/R^2,$$

where M is the mass in the sphere with radius R about the first mass point. If the motion of matter is isotropic with respect to particle one, and if both particles follow the general smoothed-out flow of matter in the region considered, M is a constant of the motion (sphere frozen into matter). For spherical symmetry of the matter distribution ρ one obtains

$$M = 4\pi \int_0^R \rho(r,t)\, r^2 dr = \text{const.}$$

For a ρ which is spatially constant one obtains in particular

$$M = 4\pi\rho R^3/3.$$

The Friedman equation appears, therefore, as the equation of the Kepler problem in the case of a vanishing angular momentum (this case is seldom treated in physics textbooks). Multiplication with $\dot{R}$ and integration gives a first integral

$$\dot{R}^2/2 - GM/R = h, \qquad h = \text{const.},$$

the energy theorem. Sometimes this equation is called the Friedman equation; h is the energy per unit mass. This equation governs also the collapse of a spherically symmetric body for vanishing pressure. From the energy integral we derive instantly

$$\dot{R}^2/R^2 = 2GM/R^3 + 2h/R^2,$$

or the following relation between Hubble constant $\dot{R}/R$ and density ρ:

$$\dot{R}/R = \pm\,(8\pi G\rho/3 + 2h/R^2)^{1/2}.$$

We obtain from the Friedman equation $\ddot{R} = -GM/R^2$ and its first integral for the Sandage constant $q = -R\ddot{R}/\dot{R}^2$

$$q = \frac{-R\ddot{R}}{\dot{R}^2} = \frac{1}{2 + 2hR/GM}.$$

For $h = 0$ the integration can be carried out immediately. One obtains

$$R(t) = (9GM/2)^{1/3}\,|t - t_0|^{2/3}.$$

This special solution for $R(t)$ is called the Einstein-deSitter solution. It gives rise to a Hubble constant

$$\frac{\dot{R}}{R} = \frac{2}{3}\frac{1}{(t - t_0)}, \qquad \frac{8\pi G\rho}{3} = \frac{4}{9}\frac{1}{(t - t_0)^2}.$$

For the Sandage constant q one has $q = \frac{1}{2}$. The integration of the Friedman equation for $h \neq 0$ becomes simple after introduction of the new time variable τ by means of $d\tau = (2|h|)^{1/2}\,dt/R(t)$. One obtains

$$h > o\colon R = \frac{GM}{2h}(\cosh\tau - 1); \qquad \pm\,(t - t_0)\,(2h)^{1/2} = \frac{GM}{2h}(\sinh\tau - \tau);$$

$$h < o\colon R = \frac{GM}{2|h|}\,(1 - \cos\tau); \qquad \pm\,(t - t_0)\,(2|h|)^{1/2} = \frac{GM}{2|h|}\,(\tau - \sin\tau).$$

The last equation corresponding to a negative constant energy per unit mass in the universe describes a cycloid. This cycloid is described by a point on the rim of a wheel with radius $GM/2|h|$ rolling on the time axis. The parameter τ is the turning angle at the hub of the wheel. The life span of the universe from creation to destruction is given by $\tau = 2\pi$ or $T = 2\pi GM/(2|h|)^{3/2}$. For the Hubble constant one gets

$$\frac{\dot{R}}{R} = \frac{2\pi}{T}\frac{\sin\tau}{(1-\cos\tau)^2}$$

and for the Sandage constant

$$q = \frac{-R\ddot{R}}{\dot{R}^2} = \frac{1}{1+\cos\tau}, \qquad \cos\tau = \frac{1}{q} - 1.$$

(For the positive energy case one has naturally $\cosh\tau = q^{-1} - 1$.) The last equation is a particularly neat relation between the dimensionless parameter τ and the dimensionless Sandage constant q. For $q = 1$, for instance, we have $\tau = 90°$. That means the distance between two particles has just reached half of the maximal value—which is obtained for $q = \infty$. We see that the Sandage constant is the most decisive parameter for the description of the universe. If q turns out to be positive (that corresponds to negative $\ddot{R}$) we have observational confirmation that gravitation still acts as an attractive force over distances of the order of a billion lightyears. If q is larger than $\frac{1}{2}$ the universe has a finite life span. If, however, $0 < q \leqq \frac{1}{2}$ the universe keeps expanding indefinitely.

It is quite clear that the derivation of the Friedman equation given above is not convincing. For a finite mass distribution with spherical symmetry it is well known that we can neglect matter outside of a sphere with radius R when we study the dynamics of this sphere. But would this still hold for infinite distributions? Another discomforting point is that the gravitational potential $-MG/R$ is centered about particle 1. If we consider a homogeneous universe particle 1 should not be distinguished. This latter point especially has led to frequent misunderstandings, e.g., the (wrong) statement that a homogeneous distribution of matter does not give rise to gravitational forces. The first point has to do with boundary conditions, the second with the fact that the Newtonian potential does not transform as a scalar under transformation from

one local inertial system to another when matter is present. The following derivation of Friedman's equation is no longer open to the objections raised above.

We postulate the Euler equations for dust with density $\rho(x_j, t)$ and velocity $V_i(x_j, t)$ in a local inertial system. The gravitational potential is given by $\phi = \phi(x_j, t)$. With $t = x_0$ we have

$$\rho_{,o} + (\rho V_i)_{,i} = 0,$$
$$V_{i,o} + V_{i,k} V_k = -\phi_{,i},$$
$$\phi_{,i,i} = 4\pi G\rho$$

These equations can easily be generalized if one wishes to account for creation of matter and momentum, a cosmologic force, or the gravitating effects of pressure. We obtain the Friedman models as those solutions of the preceding equations for which the density and velocity distributions are isotropic with respect to every dust particle, i.e.,

$$V_i(x_j, t) = \frac{\dot{R}(t)}{R(t)} x_i, \qquad \rho = \rho(t).$$

The potential ϕ is normalized by $\phi(0, t) = 0$. We get

$$M = \frac{4\pi}{3} \rho R^3 = \text{const.}, \qquad \frac{\ddot{R}}{R} = -\frac{MG}{R^2}, \qquad \phi = \frac{2\pi G\rho}{3} x_i x_i.$$

The isotropy assumption and the normalization condition for the potential fix the solution completely. We do not have to worry about the infinite amount of matter. When we transform the solution to another local inertial system we obtain again

$$\phi' = \frac{2\pi G\rho}{3} x_i' x_i',$$

the spherical symmetry of the potential about every particle of the substratum.

We shall discuss now briefly light propagation in gravitational fields. From special relativity we expect that the *local* velocity of light in every local inertial system is c. Along the path of the light ray, $x_i(t)$, one can choose now a one-parameter set of local inertial systems with parallel axes such that the local velocity of this light ray is always c. Starting from one arbitrary local inertial system at one instant one may consider all the light rays arriving

at that event. One obtains in this way a three parameter set of local inertial systems on one past light cone. This procedure is, of course, not integrable. For the Friedman models, however, it turns out that the local inertial systems obtained by starting from some arbitrary particle in the substratum lead us to the local inertial systems tied to the matter.

We have thus for an incoming light ray

$$\dot{x}_i(t) - \frac{\dot{R}}{R}x_i = -cn_i, \qquad n_i n_i = 1, \qquad n_i = \text{const.}$$

(This "photon equation" is incompatible with the equation of motion $\ddot{x}_i = -\phi_{,i}(x_j, t)$ of a free test particle; for a test particle, $R\dot{x}_i - \dot{R}x_i$ has a constant magnitude.)

In the Friedman models we have thus for an incoming light ray $r = r(t)$

$$\dot{r}(t) - \frac{\dot{R}}{R}r(t) = -c,$$

or

$$\left(\frac{r(t)}{R(t)}\right)^{\cdot} = -\frac{c}{R(t)}, \qquad \frac{r(t)}{R(t)} = -c\int_{t_1}^{t_2}\frac{d\tau}{R(\tau)}.$$

For particles of the substratum $r(t)/R(t)$ is constant. Thus, we have for signals emitted from a galaxy at times $t_1 + \Delta t_1$, received by us, respectively, at later times t_2 and $t_2 + \Delta t_2$:

$$\int_{t_1}^{t_2}\frac{d\tau}{R(\tau)} = \int_{t_1+\Delta t_1}^{t_2+\Delta t_2}\frac{d\tau}{R(\tau)},$$

or

$$\frac{\Delta t_1}{R(t_1)} = \frac{\Delta t_2}{R(t_2)}.$$

Introducing

$$z = R(t_2)/R(t_1) - 1,$$

we obtain the Doppler formula

$$\nu_1/\nu_2 = \lambda_2/\lambda_1 = 1 + z.$$

It might be worthwhile mentioning that the prescription given above for light propagation in gravitational fields leads to the

correct Pound-Rebka shift in a static gravitational field. The value for the light deflection, however, would give only half the relativistic value.

We shall conclude here our sketchy discussion of Newtonian cosmology. We saw that we obtain already the main features of relativistic cosmology with astonishing numerical coincidences. Even the law of light propagation can be visualized easily: the photon moves like a bug crawling with constant local velocity on an expanding rubber sheet. But essential features are missing; space curvature and local homogeneous Lorentz transformations. We turn now to a brief discussion of relativistic cosmology.

D. **Relativistic cosmology.** In this brief review we shall only say a few words about the Friedman models and give derivations of the Friedman differential equation. We shall not discuss inhomogeneous or anisotropic models of the universe.

For the metric we take the signature $+, -, -, -$. Greek indices run from zero to three.

The energy-momentum tensor of the matter distribution is assumed to be the one for a perfect fluid

$$T^{\mu\nu} = (\rho + p)\, U^{\mu} U^{\nu} - p g^{\mu\nu}, \qquad U^{\mu} U_{\mu} = 1.$$

The contributions to p from background radiation, cosmic rays, magnetic fields, and irregular motions of galaxies seem all to be at least more than one order of magnitude smaller than the density ρ of visible matter. It cannot be excluded, however, that $3p = \rho$ in a universe filled with neutrinos, with a negligible contribution of matter. If these neutrinos had low energies the only foreseeable mode of detection would be by their gravitational background field that would contribute to the mass of clusters. Mass determinations of some clusters of galaxies by means of the virial theorem seem to indicate the presence of large amounts of invisible matter. But nothing points specifically to neutrinos.

For the description of the present state of the universe at about 10^{10} A.D.P. (After Dense Phase) the description in terms of $p = o$ (dust) is possibly adequate. Before 10^{6} A.D.P., however, the perfect fluid with $p = \rho/3$ would seem more appropriate. We shall restrict ourselves to the discussion of the dust case. We neglect entirely the invisible matter whose density and velocity distribution is completely unknown.

1. *The Velocity Field.* We obtain from

$$T^{\mu\nu} = \rho U^\mu U^\nu \quad \text{and} \quad T^{\mu\nu}{}_{;\nu} = 0$$

that the dust is conserved and moves on geodesic world lines:

$$(\rho U^\mu)_{;\mu} = 0, \qquad U^\mu{}_{;\nu} U^\nu = 0.$$

On the Newtonian level these equations correspond to

$$\rho_{,0} + (\rho V_i)_{,i} = 0, \qquad \frac{DV_i}{Dt} = -\phi_{,i}.$$

The field of flow can be characterized locally (a differential in cosmology can be 10^9 light years long) by the first derivatives of the four vector U^μ. We define the shear tensor $q_{\mu\nu}$ by

$$q_{\nu\mu} = q_{\mu\nu} = (U_{\mu;\nu} + U_{\nu;\mu})/2 - (g_{\mu\nu} - U_\mu U_\nu) V^\lambda{}_{;\lambda}/3,$$

$$q_{\mu\nu} U^\nu = 0, \qquad q^\mu{}_\mu = 0.$$

The rotation tensor is defined by

$$-\omega_{\nu\mu} = \omega_{\mu\nu} = (U_{\mu;\nu} - U_{\nu;\mu})/2 = U_{[\mu,\nu]}, \qquad \omega_{\mu\nu} U^\nu = 0.$$

This tensor can also be expressed by a rotation vector ω^μ,

$$-\omega^\mu = \eta^{\mu\nu\alpha\beta} U_\nu \omega_{\alpha\beta}/2, \qquad \omega^\mu U_\mu = 0, \qquad \omega_{\mu\nu} = \eta_{\mu\nu\alpha\beta} U^\alpha \omega^\beta,$$

$$\omega_{\mu\nu}\omega^{\mu\nu} = -2\omega_\alpha \omega^\alpha = 2\omega^2.$$

ω is the angular velocity. The equations above follow at once from the properties of the completely skew symmetric η-tensor defined by

$$\eta^{0123} = 1, \qquad \eta^{\alpha_1 \cdots \alpha_4} \eta_{\beta_1 \cdots \beta_4} = -\det \begin{vmatrix} \delta^{\alpha_1}_{\beta_1} & \cdots & \delta^{\alpha_1}_{\beta_4} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \delta^{\alpha_4}_{\beta_1} & \cdots & \delta^{\alpha_4}_{\beta_4} \end{vmatrix}.$$

We may write now

$$U_{\mu;\nu} = q_{\mu\nu} + \omega_{\mu\nu} + \frac{1}{3}(g_{\mu\nu} - U_\mu U_\nu) U^\lambda{}_{;\lambda}.$$

In Newtonian hydrodynamics we have in complete analogy

$$V_{i,k} = q_{ik} + \omega_{ik} + \delta_{ik} V_{j,j}/3.$$

In a local inertial system with matter at rest at the origin the

relativistic quantities coincide with the Newtonian ones at the origin.

Observationally, we have some evidence that the shear tensor is small compared to the expansion scalar $V^{\lambda}{}_{;\lambda}$. An angular velocity ω of the universe larger than $V^{\lambda}{}_{;\lambda}/3$ would still be compatible with the observations. An angular velocity of $3 \times 10^{-18}\text{sec}^{-1}$ would lead to peculiar motions of galaxies of $2 \times 10^{-3}\pi/100$ years corresponding to $\omega = V^{\lambda}{}_{;\lambda}/3$. We might have to face the possibility that a rotation of the universe is dynamically a more important feature than its expansion.

There exists, however, a serious difficulty in the discussion of these observations. We cannot actually see the field of flow in the universe as it is or was. The curved space-time manifold is an extremely complicated and essentially unknown optical system that might distort quite considerably the picture that we receive from distant galaxies. Intervening large scale gravitational waves would make the world appear quite different from what it actually is. J. Kristian and R. K. Sachs were the first to discuss this question in some detail.

In the following we shall assume that rotation and shear vanish.

We see that already on the kinematical level, before we can ask any questions of a dynamical nature, the nature of the universe is still poorly understood.

2. *Dynamics.* We start with a few definitions. For a vector field U_{α} we have

$$U_{\alpha;\beta;\chi} - U_{\alpha;\chi;\beta} = -R^{\delta}{}_{\cdot\alpha\beta\chi}\,U_{\delta}, \qquad -R_{\alpha\beta} = R^{\chi}{}_{\cdot\alpha\beta\chi},$$

$$R = R^{\alpha}{}_{\cdot\alpha}, \quad -C_{\alpha\beta} = -\frac{1}{2}R_{\alpha\beta} + \frac{1}{12}g_{\alpha\beta}R, \quad G_{\alpha\beta} = R_{\alpha\beta} - \frac{1}{2}g_{\alpha\beta}R.$$

The Weyl conformal curvature tensor is defined by

$$C_{\alpha\beta\chi}{}^{\delta}{}_{\cdot} = R_{\alpha\beta\chi}{}^{\delta}{}_{\cdot} + \delta^{\delta}_{\alpha}C_{\beta\chi} - \delta^{\delta}_{\beta}C_{\alpha\chi} + C_{\alpha}{}^{\delta}{}_{\cdot}g_{\beta\chi} - C_{\beta}{}^{\delta}{}_{\cdot}g_{\alpha\chi}.$$

A vanishing Weyl tensor allows us to express the Riemann tensor $R_{\alpha\beta\chi\delta}$ in terms of the Ricci tensor. We obtain

$$R_{\alpha\beta\chi}{}^{\delta}{}_{\cdot} = -\frac{1}{6}(g_{\alpha\chi}\delta^{\delta}_{\beta} - g_{\beta\chi}\delta^{\delta}_{\alpha})\,R - \frac{1}{2}(\delta^{\delta}_{\alpha}R_{\beta\chi} - \delta^{\delta}_{\beta}R_{\alpha\chi} + g_{\beta\chi}R^{\delta}_{\alpha} - g_{\alpha\chi}R^{\delta}_{\beta}).$$

We use the field equations in the form

$$G_{\alpha\beta} + \Lambda g_{\alpha\beta} = -\kappa T_{\alpha\beta}, \qquad \kappa = 8\pi G/c^2.$$

where Λ is the cosmologic constant. Equivalently we obtain

$$R_{\alpha\beta} = \left(\Lambda + \kappa\frac{T}{2}\right) g_{\alpha\beta} - \kappa T_{\alpha\beta}, \qquad R = 4\Lambda + \kappa T.$$

For a conformally flat metric ($C_{\alpha\beta\chi}{}^{\delta} = 0$) we can express the Riemann tensor via field equations by Λ and the energy momentum tensor. We obtain

$$R_{\alpha\beta\chi}{}^{\delta} = -\frac{1}{3}(g_{\beta\chi}\delta^{\delta}_{\alpha} - g_{\alpha\chi}\delta^{\delta}_{\beta})(\Lambda + \kappa T)$$

$$+\frac{\kappa}{2}(\delta^{\delta}_{\alpha}T_{\beta\chi} - \delta^{\delta}_{\beta}T_{\alpha\chi} + g_{\beta\chi}T^{\delta}_{\alpha} - g_{\alpha\chi}T^{\delta}_{\beta}).$$

We consider now different characterizations of the Friedman models and different derivations of the Friedman equation. (In literature the Friedman models are often called after Robertson and Walker.)

(a). *Characterization by the field of flow (dust).* We assume first that shear and rotation vanish along the world line of one dust particle:

$$q_{\mu\nu} = \omega_{\mu\nu} = 0,$$

$$U_{\mu;\nu} = \frac{1}{3}(g_{\mu\nu} - U_{\mu}U_{\nu})\,U^{\lambda}{}_{;\lambda}.$$

We have then for the derivative of $U^{\lambda}{}_{;\lambda}$ along the world line

$$\begin{aligned}\frac{D}{Ds}U^{\lambda}{}_{;\lambda} &\equiv U^{\lambda}{}_{;\lambda;\mu}U^{\mu}\\ &= (U^{\lambda}{}_{;\lambda;\mu} - U^{\lambda}{}_{;\mu;\lambda})\,U^{\mu} + U^{\lambda}{}_{;\mu;\lambda}U^{\mu}\\ &= -R^{\delta\lambda}{}_{\cdot\cdot\,\lambda\mu}U_{\delta}U^{\mu} + (U^{\lambda}{}_{;\mu}U^{\mu})_{;\lambda} - U^{\lambda}{}_{;\mu}U^{\mu}{}_{;\lambda}\\ &= +R_{\delta\mu}U^{\delta}U^{\mu} - \left(\frac{1}{3}U^{\lambda}{}_{;\lambda}\right)^2(g_{\lambda\mu} - U_{\lambda}U_{\mu})(g^{\lambda\mu} - U^{\lambda}U^{\mu})\\ &= \left(\Lambda + \frac{\kappa T}{2}\right) - \kappa T_{\delta\mu}U^{\delta}U^{\mu} - 3\left(\frac{1}{3}U^{\lambda}{}_{;\lambda}\right)^2.\end{aligned}$$

With $T_{\delta\mu} = \rho U_{\delta}U_{\mu}$ we get

$$\frac{D}{Ds}\left(\frac{1}{3}\,U^{\lambda}{}_{;\lambda}\right) - \left(\frac{1}{3}\,U^{\lambda}{}_{;\lambda}\right)^2 = \frac{\Lambda}{3} - \frac{\kappa\rho}{6}.$$

We introduce now a scalar R by the equation

$$\frac{1}{3}\,U^{\lambda}{}_{;\lambda} = \frac{D \ln R}{Ds} = \frac{1}{R}\frac{DR}{Ds}.$$

The continuity equation $(\rho U^{\mu})_{;\mu} = 0$ gives then

$$\frac{4\pi}{3}\,\rho R^3 = M = \text{const. along world line.}$$

We obtain

$$\frac{D^2R}{Ds^2} = \frac{\Lambda}{3}\,R - \frac{G}{c^2}\frac{M}{R^2},$$

the local Friedman equation along one world line (by keeping rotation and shear we would have obtained the Raychaudhuri equation). This is probably the shortest relativistic derivation of the Friedman equation.

The Friedman models can then be obtained from the postulate that shear and rotation vanish along every world line of matter.

(b). *Characterization through conformal flatness.* The simplest nontrivial space-time metrics are those which are conformally flat. This is locally expressed by the vanishing of the Weyl tensor. The intrinsic degrees of freedom of Einstein's gravitational field are not excited in that case.

We obtain the Friedman equation for those metrics by studying the geodesic deviation.

Consider two neighboring (time-like) geodesics with tangent vectors belonging to the vector field U^{μ}. We introduce a space-like connecting vector η^{μ} orthogonal on the first geodesic,

$$U^{\mu}\eta_{\mu} = 0.$$

One obtains then for the second derivative of η_{μ} along the first geodesic

$$\frac{D^2\eta_{\alpha}}{Ds^2} = -\,R_{\alpha\beta\chi\cdot}{}^{\delta}\,U^{\beta}U^{\chi}\eta_{\delta}.$$

This is the equation of geodesic deviation. In Newtonian theory

we have in a local inertial system where ϕ and its first derivatives vanish in the origin

$$\ddot{\eta}_i = -\phi_{,i,k}\eta_k$$

as the first terms in a series development of the potential.

Before we proceed to the derivation of the Friedman equation we consider a case slightly more general than necessary. We assume conformal flatness and take for the energy-momentum tensor one of a perfect fluid with velocity vector V^μ, which need not coincide with the tangent vector U^μ,

$$T_{\alpha\beta} = (\rho + p)\, V_\alpha V_\beta - p g_{\alpha\beta}, \qquad T = \rho - 3p, \qquad V_\alpha V^\alpha = 1.$$

We express the Riemann tensor by the energy momentum tensor in the equation for geodesic deviation and obtain

$$\frac{D^2\eta_\alpha}{Ds^2} = \left[\frac{1}{3}(\Lambda + \kappa\rho) - \frac{\kappa(\rho+p)}{2}(U_\beta V^\beta)^2\right]\eta_\alpha + \frac{\kappa(\rho+p)}{2}(U_\alpha U_\beta V^\beta - V_\alpha)\, V^\delta \eta_\delta.$$

This formula shows us that gravitational accelerations in particles that move extremely relativistic with respect to the universe become arbitrarily strong. Orthogonal to its direction of motion the particle is squeezed together by an acceleration gradient of

$$-\frac{4\pi G(\rho+p)}{1-\beta^2}\ \mathrm{sec}^{-2}, \qquad \beta \to 1,$$

as measured in its rest system. I do not know if this effect could be important under any conditions in nature (in the Schwarzschild field one gets a similar effect that leads at high enough energies to ionization of high energy atoms and fission of nuclei at still higher energies).

We leave now the field of gravitational fiction in a hurry and put $V_\alpha = U_\alpha$. We then obtain

$$\frac{D^2\eta_\alpha}{Ds^2} = \left[\frac{\Lambda}{3} - \frac{\kappa(\rho+3p)}{6}\right]\eta_\alpha.$$

This is again the Friedman equation with additional terms for cosmological constant and pressure. One sees that pressure in-

creases the gravitational attraction. The Poisson equation of Newtonian theory has to be replaced by

$$\nabla^2\phi + \Lambda c^2 = 4\pi G(\rho + 3p/c^2).$$

The equation of geodesic deviation shows that the gravitational field of conformally flat metrics is isotropic in the rest system of matter.

Using the full system of field equations one can show that space-time has a Friedman metric (Trumper's theorem). The converse is also true: All Friedman metrics are conformally flat. This last remark simplifies considerably the discussion of light propagation in Friedman models.

(c). *Group theoretical characterization.* The Friedman models have a metric

$$ds^2 = dt^2 - R^2(t)\, d\sigma^2,$$

where $d\sigma$ is the line element of a three-dimensional space of constant curvature. Such spaces are homogeneous and isotropic with six-dimensional groups of motions. One obtains the Friedman equation of Newtonian theory by putting $|h| = c^2$. For negative energy per mass one obtains positive curvature for the three-space, and for positive energy negative curvature. The discussion of the Friedman models based on the metric above can be found in most textbooks and will not be repeated here.

3. *Space-Time Diagram.* Space-time diagrams are often used in special relativity. For cosmology such maps are very helpful for obtaining an insight into the qualitative features of different world models. Due to the isotropy of the Friedman models, a one-dimensional space substitutes well for the three-dimensional one.

Taking X as a radial coordinate in the three-space we study the two-dimensional manifold with metric

$$ds^2 = dt^2 - R^2(t)\, dX^2.$$

The world lines of matter are given by $X = \text{const.}$ $R(t)$ will be taken as a solution of the Friedman equation.

The first map one might think of is the conformal one,

$$ds^2 = R^2(t')\,[dt'^2 - dX^2], \qquad dt' = dt/R(t).$$

In this map light propagation is the same as in Minkowski

space. The neighborhood of the singularity is, however, badly distorted. The time parameter t' is essentially equal to τ that appeared in the parameter representation of the Friedman solutions. The reader can easily construct such maps—with z-values of quasars inserted—for different q-values.

I shall discuss here another map that gives a more intuitive picture of the singularity. I introduce a new x-coordinate which is the instantaneous distance measured along a spatial geodesic in the subspace $t = \text{const}$. We have

$$x = R(t)\,X, \qquad dx = \frac{\dot{R}}{R}dt \cdot x + R(t)\,dx.$$

We obtain, therefore,

$$ds^2 = dt^2\left[1 - \left(\frac{\dot{R}}{R}x\right)^2\right] + \frac{\dot{R}}{R}x\,dx\,dt - dx^2.$$

The world lines of the galaxies are now given by $x = R(t)\,X$ with constant X. The map is undistorted along our world line given by $x = 0$. The light cone can be easily expressed in terms of the parameter τ of the Friedman solutions.

This map shows an interesting feature of the Friedman solutions: When we go back on the light cone into the past the distance x of galaxies, visible now, increases, reaches a maximum and decreases towards the singularity. This maximum occurs, for instance, for $q = 1/2$ at a red shift $z = 5/4$.

I did not find the time to review here the present status of observational cosmology. Many interesting theoretical developments have not even been mentioned. I apologize for these omissions.

UNIVERSITY OF TEXAS
AUSTIN, TEXAS

David Layzer

A Unified Approach to Cosmology

I. **Introduction.** Cosmology is the branch of astronomy that deals with the structure and evolution of the universe as a whole. Since only a limited portion of the universe is accessible to observation, cosmological assumptions can never be verified directly. In this respect cosmological theories resemble other astronomical theories, such as the theory of stellar evolution, whose underlying assumptions are linked to its verifiable predictions by a long and complex chain of inference. Cosmology does have one distinctive characteristic, however. Other branches of astronomy treat classes of similar but not identical objects; cosmology treats a unique system.

The uniqueness of the universe raises a question that does not arise elsewhere in astronomy. Are the defining properties of the universe (the subject of cosmological hypotheses) to be regarded as accidental or as essential? That is, are they to be regarded as analogous to the defining properties of a particular star or to those of a fundamental particle? In the first case they would be merely elements in a statistical description, insignificant in themselves; in the second case they could be regarded as laws of nature.

These lectures proceed from the view that the defining properties

of the universe have the status of natural laws. Whether or not this approach is correct, it has the advantage of directing us toward the simplest hypotheses. Thus it encourages us to construct theories that are especially vulnerable to observational disproof. In these lectures I shall develop some consequences of what appear to me to be the simplest cosmological postulates that are consistent with currently accepted physical laws.

II. **The principle of uniformity.** Astronomical observations have shown that the structure of the observable universe is exceedingly complex. Matter is distributed in a nested sequence, or hierarchy, of self-gravitating systems whose linear dimensions range from a few tens of kilometers to one hundred million light years and whose mean densities range from a few atoms per cubic meter to a million times the density of water. In the nineteenth century astronomers conjectured that the hierarchy might extend indefinitely toward systems with larger dimensions and smaller mean densities. Hubble and his successors showed, however, that it eventually terminates. The largest astronomical systems are clusters, or perhaps superclusters, of galaxies.

Although the diameter of a galaxy cluster is large by ordinary astronomical standards, it is less than one percent of the radius of the observable universe. In principle, we may therefore compare the statistical properties of nonoverlapping volumes of space whose dimensions are large compared with the scale of the local structure. In practice such comparisons are complicated by errors of measurement that depend systematically on distance. They are further complicated by the fact that astronomical observations provide information only about events on the observer's light cone, so that observations of events that are remote in space refer to epochs that are remote in time. In spite of these difficulties, observational evidence gathered during the last 40 years strongly supports the hypothesis that all "fair samples" of the observable universe, considered at a common epoch, have the same statistical properties; that is, no average property of the observed distribution of galaxies serves to single out a preferred position or direction in space. This conclusion was first enunciated by Hubble, who called it the "principle of uniformity."

By an "average" property of the observed distribution is meant a spatial average extended over a region whose dimensions are

large compared with the scale of the local structure. Suppose for the sake of definiteness that this region is a sphere of radius R. The average value of a quantity pertaining to the distribution of galaxies cannot be strictly independent of R, but may approach a definite limit as $R \to \infty$. On the other hand, R must remain finite if we are to compare the average properties of distinct regions. In practice this requirement presents no difficulty, because there exist values of R that are both large compared with the scale of the local structure and small compared with the radius of the observable universe. But a precise mathematical formulation of the principle of uniformity is impossible unless we attach a more sophisticated interpretation to the term "average property."

We shall make use of a mathematical description of the kind employed in modern theories of turbulence, where the density and velocity fields are represented by random functions and where spatial averages are replaced by ensemble averages. We shall postulate that the cosmic density and velocity fields are statistically homogeneous and isotropic. The additional assumption that the local structure has a finite scale ensures, according to the ergodic theorems of Birkhoff and Khinchin, that spatial averages converge to the corresponding ensemble averages. Thus Hubble's principle of uniformity emerges as a consequence of our postulates.

III. **The substratum: relativistic cosmology.** Our first objective is to develop a dynamical theory applicable to a statistically homogeneous and isotropic distribution of matter. We shall find that first-order perturbation theory is adequate for this purpose. In this section we consider the "unperturbed" distribution, or substratum. The substratum is described by the classical cosmological theory of Einstein and Friedman, whose main features we now review briefly.

It is well known that the line element for a distribution of matter that is everywhere isotropic can be written in the form[1]

$$ds^2 = -g_{ij}dx^i dx^j = c^2dt^2 - S^2(t)\,d\bar{l}^2, \tag{3.1}$$

where t is the local proper time (as well as the "cosmic" time) and $d\bar{l}$ is the element of length in a space of constant curvature,

[1] Latin indices run from 0 to 3, Greek indices from 1 to 3, and a summation is implied over repeated indices; $x^0 = ct$.

given explicitly by

$$d\bar{l}^2 = d\bar{r}^2/(1 - k\bar{r}^2) + \bar{r}^2(d\bar{\theta}^2 + \sin^2\bar{\theta}\, d\bar{\phi}^2), \tag{3.2}$$

where $k = +1$, 0, or -1, according as the curvature of space is positive, 0, or negative. The coordinates $\bar{r}$, $\bar{\theta}$, $\bar{\phi}$ are comoving; an observer with fixed coordinates ($\bar{r}$, $\bar{\theta}$, $\bar{\phi}$) is locally at rest and sees an isotropic distribution of matter and motion. Hence the nonvanishing components of the energy-momentum tensor must have the form

$$\bar{T}^0_0 = -\epsilon \equiv -\bar{\rho}c^2, \qquad \bar{T}^1_1 = \bar{T}^2_2 = \bar{T}^3_3 = \bar{p}. \tag{3.3}$$

Einstein's field equations

$$R_{ik} = \frac{8\pi G}{c^4} T^*_{ik} \equiv \frac{8\pi G}{c^4}\left(T_{ik} - \frac{1}{2} g_{ik} T\right) \tag{3.4}$$

provide two independent relations connecting the functions S, ϵ, and $\bar{p}$:

$$\frac{1}{2}\dot{S}^2 - \frac{4\pi}{3} G\bar{\rho}S^2 = -\frac{1}{2}kc^2, \tag{3.5}$$

$$d(\epsilon S^3)/dt + \bar{p}\, dS^3/dt = 0. \tag{3.6}$$

These equations need to be supplemented by a third relation between ϵ, $\bar{p}$, and S. An equation of state for the substratum would furnish such a relation. Let us consider some examples.

(a) *Perfect ultrarelativistic gas.* The equation of state is

$$\epsilon = 3\bar{p} \tag{3.7}$$

whence, by (3.6),

$$\epsilon S^4 = \text{const.} \tag{3.8}$$

With the help of this relation we can integrate Equation (3.5). For sufficiently small values of S, the two terms on the left side of (3.5) are individually large, and we have the asymptotic solution

$$S \sim t^{1/2} \tag{3.9}$$

which is valid for sufficiently small values of t when $k = \pm 1$, and for all values of t when $k = 0$.

(b) *Perfect nonrelativistic gas.* The equation of state is

$$\epsilon = \bar{\mu}c^2 + 3\bar{p}/2 \tag{3.10}$$

where $\bar{\mu}$ denotes the rest-mass density. In this case $\bar{p} \ll \bar{\mu}c^2$, and it is therefore a good approximation to set $\bar{p} = 0$. From Equation (3.6) we then obtain

$$\epsilon S^3 = \text{const.}, \tag{3.11}$$

and integration of (3.5) yields

$$\begin{aligned} S &= k^{-1}S_0(1 - \cos k^{1/2}u), \\ ct &= k^{-1}S_0(u - k^{-1/2}\sin k^{1/2}u). \end{aligned} \tag{3.12}$$

For sufficiently small values of t the solutions for $k = \pm 1$ merge with the solution $k = 0$:

$$S \sim t^{2/3}, \qquad 6\pi G\bar{p}t^2 = 1, \qquad (k = \bar{p} = 0). \tag{3.13}$$

In all cases the substratum begins to expand from a singular initial state of infinite density. Models with negative or zero spatial curvature ($k \leqq 0$), which can have infinite volume, expand indefinitely (or contract from an infinitely dispersed initial state in the infinitely remote past). Models with positive spatial curvature ($k > 0$), which have finite volume, expand to a state of minimum density and then contract along the time-reversed path.

(c) *Imperfect nonrelativistic gas.* The equation of state has the form

$$\epsilon = \bar{\mu}c^2 + \epsilon_k + \epsilon_i \tag{3.14}$$

where ϵ_k denotes the kinetic-energy density and ϵ_i denotes the interaction-energy density. The virial theorem provides a further relation:

$$3\bar{p} = 2\epsilon_k + \langle \mathbf{x} \cdot \mathbf{F} \rangle \tag{3.15}$$

where $\mathbf{F}$ denotes the intermolecular force. When the form of $\mathbf{F}$ has been specified, both the interaction-energy density ϵ_i and the virial $\langle \mathbf{x} \cdot \mathbf{F} \rangle$ can be calculated. Thus Equations (3.14), (3.15), (3.5), and (3.6) form a complete set.

We shall find that the equation of state for a nonrelativistic imperfect gas adequately describes the substratum during all but the earliest stages of its history. Short-range gravitational forces associated with the fluctuating density field play a role analogous to that of intermolecular forces in an ordinary gas.

IV. **Fluctuations: the Newtonian approximation.** Turning now to the fluctuation part of the distribution, we may write

$$g_{ij} = \bar{g}_{ij} + \tilde{g}_{ij}, \quad R_{ij} = \bar{R}_{ij} + \tilde{R}_{ij}, \quad T_{ij} = \bar{T}_{ij} + \tilde{T}_{ij}, \tag{4.1}$$

where the quantities distinguished by an overbar refer to the substratum. From Einstein's field Equations (3.4) we obtain

$$\tilde{R}_{ij} = (8\pi G/c^4)\, \tilde{T}^*_{ij}. \tag{4.2}$$

The tensor $\tilde{R}_{ij}$ is a nonlinear function of the $\tilde{g}_{ij}$. However, if the fluctuating gravitational fields are sufficiently weak, we may neglect contributions to $\tilde{R}_{ij}$ that are quadratic in the $\tilde{g}_{ij}$. It is important to notice that this approximation does not require the fluctuating component of the density to be small compared with the mean density. Even under present conditions, when the density fluctuates by a factor of 10^{30} or more between stellar interiors and intergalactic space, the fluctuating gravitational field is weak enough to justify the linear approximation except, possibly, in the interiors of highly collapsed bodies—which, in any case, can be treated separately, since their interaction with the rest of the universe is negligible.

The $\tilde{g}_{ij}$ are determined by (4.2) only up to an infinitesimal coordinate transformation. As Einstein emphasized in his 1916 paper, the resulting freedom in the choice of coordinates can be exploited to simplify the form of the field equations. Harmonic coordinates are defined by the four conditions

$$\tilde{g}^i_{j;i} - \tilde{g}^i_{i;j}/2 = 0. \tag{4.3}$$

When these conditions are imposed, the linearized field Equations (4.2) take the form (Lanczos [5])

$$\tilde{g}_{ij;k}^{;k} - \bar{g}^{kr}(\bar{R}_{irjk} + \bar{R}_{jrki}) = (16\pi G/c^4)\, \tilde{T}^*_{ij}. \tag{4.4}$$

In (4.3) and (4.4) the $\tilde{g}_{ij}$ represent tensors in the space-time of the substratum.

Let λ denote a characteristic length associated with spatial variations of the fluctuating gravitational field. We now assume that

$$\lambda \ll c\dot{S}/S \equiv cH, \qquad \lambda \ll S \qquad (k = \pm 1). \tag{4.5}$$

The first inequality states that the scale of the fluctuating field is much less than the radius of the observable universe, the second that it is much less than the radius of curvature of the universe. Under these assumptions, the second term on the left side of (4.4) is negligible compared with the first term, and the equations take

the simple, uncoupled form

$$\tilde{g}^{i;k}_{j;k} = (16\pi G/c^4)\,\widetilde{T}^{*i}_{j}. \tag{4.6}$$

So far we have assumed that the fluctuating gravitational field is weak and that its characteristic scale is small compared with the radius of the observable universe and with the radius of space curvature. We now assume further that the characteristic velocities associated with the fluctuating gravitational field and with the fluctuating density field that gives rise to it are small compared with the speed of light. We may then neglect time-derivatives of the $\tilde{g}^i_j$ on the left side of Equation (4.6); and in the formula

$$T^i_j = -(\epsilon + p)\,u^i u_j + \delta^i_j p, \tag{4.7}$$

for the energy-momentum tensor of a perfect fluid we may set $u_i = \delta_{0i}$. The 00-component of (4.6) then reduces to

$$\tilde{g}^{0;\alpha}_{0;\alpha} = -(8\pi G/c^2)\,(\tilde{\rho} + 3\tilde{p}/c^2). \tag{4.8}$$

Setting

$$\tilde{g}^0_0 = -2\phi/c^2 \tag{4.9}$$

and assuming that $|\tilde{p}| \ll |\tilde{\rho}|\,c^2$, we finally obtain

$$(1/S^2)\,(\partial^2\phi/\partial\bar{x}^\alpha\partial\bar{x}^\alpha) \equiv \partial^2\phi/\partial x^\alpha\partial x^\alpha = 4\pi G\tilde{\rho}. \tag{4.10}$$

Here we have introduced the locally Galilean coordinates

$$x^\alpha = S(t)\,\bar{x}^\alpha. \tag{4.11}$$

We regard the density field ρ at a given instant of cosmic time as a statistically homogeneous random function whose mathematical expectation is the density $\bar{\rho}$ of the substratum:

$$\langle\rho\rangle = \bar{\rho}, \qquad \langle\tilde{\rho}\rangle \equiv \langle(\rho - \bar{\rho})\rangle = 0. \tag{4.12}$$

The potential ϕ is determined by Equation (4.10) up to an additive constant, which we choose so as to make the mathematical expectation of ϕ vanish. Then ϕ is given by the familiar formula

$$\phi(\mathbf{x},t) = -G\int\frac{\tilde{\rho}(\mathbf{x}-\mathbf{x}')\,d^3\mathbf{x}'}{|\mathbf{x}-\mathbf{x}'|}. \tag{4.13}$$

To ensure the convergence of the integral on the right we would need to impose somewhat restrictive, though not entirely artificial,

conditions on the fluctuating density field. We shall not do this, because ϕ itself has a purely formal significance. The measurable quantities that can be formed from ϕ, such as the fluctuating gravitational force and the potential energy per unit mass stored in the fluctuations, are expressed by integrals that converge under much less restrictive conditions.

Let $\mathbf{v}$ denote the velocity of a particle relative to the substratum. Then

$$v^\alpha = S(t)\, d\bar{x}^\alpha/dt. \tag{4.14}$$

It is important to distinguish between the peculiar velocity $\mathbf{v}$ and the velocity $\mathbf{V}$ measured in a locally Galilean frame of reference. The latter is given by

$$V^\alpha = dx^\alpha/dt = d(S\bar{x}^\alpha)/dt = v^\alpha + Hx^\alpha. \tag{4.15}$$

It can be shown that, under the approximations made above, the equations of motion for a test particle take the form

$$d(Sv^\alpha)/dt = -S\,\partial\phi/\partial x^\alpha \tag{4.16a}$$

or equivalently,

$$dv^\alpha/dt = -Hv^\alpha - \partial\phi/\partial x^\alpha. \tag{4.16b}$$

Equations (4.16a, b) express the equations of motion in a form that makes no reference to any fixed point in space. An alternative form that refers to a fixed locally Galilean frame of reference results from differentiating (4.15):

$$dV^\alpha/dt = dv^\alpha/dt + \dot{H}x^\alpha + HV^\alpha = -\partial\phi/\partial x^\alpha + (\dot{H} + H^2)x^\alpha. \tag{4.17}$$

From Equations (3.5) and (3.6) we obtain

$$\dot{H} + H^2 = \ddot{S}/S = -(4\pi G/3)\,(\bar{\rho} + 3\bar{p}/c^2), \tag{4.18}$$

whence

$$dV^\alpha/dt = -\partial\phi/\partial x^\alpha - (4\pi G/3)\,(\bar{\rho} + 3\bar{p}/c^2)\,x^\alpha. \tag{4.19}$$

These equations have the following simple interpretation. The acceleration of a test particle may be set equal to the sum of two Newtonian forces. The first is derived from the potential ϕ and results from the fluctuating component of the matter distribution. The second can be thought of as arising from a finite, uniform

sphere of density $(\bar{\rho} + 3\bar{p}/c^2)$ centered on the origin and including the particle under consideration.

Equations (4.16a, b) show that in a universe without irregularities ($\phi \equiv 0$) the direction of motion of a test particle remains constant while its speed (relative to the substratum) diminishes as S^{-1}. This is a special case of a more general result that applies at relativistic as well as nonrelativistic particle energies. The equations of motion for a free particle show that the direction of motion remains constant while the speed diminishes according to the law

$$v/(1 - v^2/c^2)^{1/2} \propto S^{-1}; \tag{4.20}$$

that is, the momentum of a free test particle diminishes as S^{-1}. This result is valid also for particles of zero rest mass.

Equation (4.20) can be derived without the use of general relativity. A test particle moving in a uniform, isotropic universe always sees a perfectly isotropic distribution of matter. By symmetry, the acceleration experienced by a particle must, therefore, be kinematic in origin; that is, it must result from the expansion of the substratum and the definition of v^α as a velocity relative to the substratum. On using the special-relativistic law for the composition of velocities to calculate the change of v resulting from the expansion of the substratum, one easily obtains (4.20).

The preceding discussion of the Newtonian approximation closely parallels Einstein's discussion, in the 1916 paper, of the Newtonian approximation for an island universe embedded in empty space. Apart from the fact that the assumption of finite density for the substratum accords better with observation than the assumption of zero density, the present treatment has one important advantage over Einstein's original discussion. In a nearly empty universe, no property of the distribution of matter serves to distinguish "nonrotating" coordinate systems (Galilean at infinity) from "rotating" coordinate systems (non-Galilean at infinity). In a statistically homogeneous and isotropic universe, on the other hand, coordinate systems in which the equations of motion assume the Newtonian form are, as we have seen, related in a definite way to the cosmic coordinate system in which the postulated symmetry properties of the distribution find their simplest expression. In short, the postulate of statistical homogeneity and isotropy ensures that the theory satisfies the famous requirement laid down by Mach,

that the locally inertial coordinate system should be determined by the distribution of matter and motion on a cosmic scale. This advantage of homogeneous and isotropic world-models seems to have been clearly recognized by Einstein; see [3].

V. **The energy equation.** As in Newtonian mechanics, one can derive an energy-balance equation from the equations of motion. I shall here summarize the results; derivations may be found in [7] and [8].

We consider a distribution consisting of gravitating particles. The following definitions arise naturally in the course of the derivation. The mean kinetic energy per unit mass associated with the fluctuating component of the distribution is

$$T_m \equiv \frac{1}{2}\langle v^2\rangle_m = \lim_{V\uparrow} (\bar{\rho}V)^{-1}\int_V \frac{1}{2}\rho v^2 dV, \tag{5.1}$$

where the notation $\langle\ \rangle_m$ indicates an ensemble average with mass-density as the weighting function. The equality between the ensemble average and the limit of the space average is a consequence of the ergodicity of the distribution, which in turn follows from the assumption that the local irregularities have a finite scale.

The mean potential energy per unit mass is defined by

$$U_m \equiv \frac{1}{2}\langle\phi\rangle_m = \lim_{V\uparrow} (\bar{\rho}V)^{-1}\int \frac{1}{2}\rho\phi dV. \tag{5.2}$$

On inserting formula (4.13) for ϕ into this definition we obtain

$$\begin{aligned}\frac{1}{2}\langle\phi\rangle_m &= \frac{1}{2}\bar{\rho}^{-1}\langle\rho\phi\rangle = \frac{1}{2}\bar{\rho}^{-1}\langle\tilde{\rho}\phi\rangle \\ &= \frac{1}{2}\bar{\rho}^{-1}\left\langle -G\int \frac{\tilde{\rho}(\mathbf{x},t)\,\tilde{\rho}(\mathbf{x}',t)\,d^3\mathbf{x}'}{|\mathbf{x}-\mathbf{x}'|}\right\rangle \\ &= -\frac{1}{2}G\bar{\rho}^{-1}\int \frac{\langle\tilde{\rho}(\mathbf{x},t)\,\tilde{\rho}(\mathbf{x}',t)\rangle\,d^3x'}{|\mathbf{x}-\mathbf{x}'|}\end{aligned} \tag{5.3}$$

where the brackets $\langle\ \rangle$ indicate an unweighted ensemble average. The assumption of statistical homogeneity and isotropy implies that the covariance of the density depends only on the scalar quantity $|\mathbf{x}-\mathbf{x}'|$. We may therefore write

$$\langle\tilde{\rho}(\mathbf{x},t)\,\tilde{\rho}(\mathbf{x}',t)\rangle = \langle\tilde{\rho}^2\rangle f(|\mathbf{x}-\mathbf{x}'|), \tag{5.4}$$

where the dimensionless autocorrelation function $f(r)$ assumes the value 1 at $r = 0$. Hence (5.3) becomes

$$\langle \phi \rangle_m / 2 = -2\pi G \bar{\rho} \lambda^2 \alpha^2, \tag{5.5}$$

where the clustering scale λ and the clustering amplitude α are defined by

$$\lambda^2 = \int_0^\infty f(r)\, r dr \tag{5.6}$$

and

$$\langle \tilde{\rho}^2 \rangle = \alpha^2 \bar{\rho}^2. \tag{5.7}$$

Formula (5.5) shows that the mean potential energy per unit mass is negative definite, an important result.

We have already seen that in the absence of density fluctuations, the peculiar velocity of a test particle decreases as the universe expands. Thus even in this simple case energy is not conserved. When density fluctuations are present, the energy-balance equation is found to take the form

$$dE_m/dt \equiv d(T_m + U_m)/dt = -H(2T_m + U_m). \tag{5.8}$$

Although the definitions of T_m and U_m are reminiscent of ordinary Newtonian mechanics, there are several important differences between the energy balance of an isolated Newtonian system and that of local irregularities in a statistically homogeneous and isotropic universe. (a) T_m and U_m are associated with the fluctuating component of the distribution only. It is not possible to define corresponding quantities for the substratum, since the kinetic and potential energies associated with a spherical volume of radius R vary as R^5. (b) The potential energy of an isolated system is not an additive quantity, as is the potential energy associated with fluctuations in a statistically homogeneous and isotropic universe. The additive character of the potential energy for a cosmic distribution depends on the fact that the peculiar density $\tilde{\rho}$ assumes both positive and negative values. The fluctuating gravitational forces derived from the potential ϕ are in fact short-range forces, like the screened Coulomb forces in a plasma. The characteristic length λ defined by Equation (5.6) measures the range of these forces. (c) As Equation (5.8) shows, the sum of the kinetic and potential

energies per unit mass is not in general constant in time. The right side of the energy-balance equation has a simple kinematic interpretation. As we have already noted, the velocity of a freely moving particle varies as S^{-1}; this effect is responsible for the first term on the right side of (5.8). The second term is a consequence of the expanding length scale and the fact that the potential energy per unit mass is a homogeneous function of degree -1 in the mutual separations of the particles that make up the distribution.

Since the fluctuating gravitational forces have finite range, we may regard the distribution of gravitating particles as an imperfect gas. Comparison between the energy-balance Equation (5.8) and the cosmological relation (3.6) shows that

$$\epsilon = \bar{\rho}(T_m + U_m) = \langle \rho v^2 \rangle / 2 + \langle \rho \phi \rangle / 2 \tag{5.9}$$

and

$$3\bar{p} = \bar{\rho}(2T_m + U_m) = \langle \rho v^2 \rangle + \langle \rho \phi \rangle / 2. \tag{5.10}$$

The last equation has precisely the form predicted by elementary kinetic theory, namely, Equation (3.15), since in the case of inverse-square attractive forces

$$\langle \mathbf{x} \cdot \mathbf{F} \rangle = \langle U \rangle. \tag{5.11}$$

VI. **The virial theorem.** The kinetic and potential energies of a system of gravitating particles in equilibrium satisfy the relation

$$2T + U = 0. \tag{6.1}$$

One derives this relation from the equations of motion $\dot{\mathbf{p}} = \mathbf{F}$ by taking the scalar product of both sides with $\mathbf{x}$ and summing (or averaging) over the system:

$$\langle \mathbf{x} \cdot \dot{\mathbf{p}} \rangle = d\langle \mathbf{x} \cdot \mathbf{p} \rangle / dt - \langle \mathbf{v} \cdot \mathbf{p} \rangle = \langle \mathbf{x} \cdot \mathbf{F} \rangle. \tag{6.2}$$

If the system is in equilibrium and consists of a large number of particles, the time-derivative vanishes, and we obtain

$$\langle \mathbf{v} \cdot \mathbf{p} \rangle + \langle \mathbf{x} \cdot \mathbf{F} \rangle = 0, \tag{6.3}$$

which reduces to (6.1) for nonrelativistic particle-velocities and inverse-square interparticle forces. In Equation (6.2) the brackets $\langle\ \rangle$ may also be interpreted as indicating a time average for a single particle; the time-derivative then vanishes if the motion is periodic or quasi-periodic.

In attempting to extend these considerations to a cosmic distribution of gravitating particles, one encounters two main difficulties. (a) In general a particle will not remain within any fixed comoving region of finite extent, nor will a group of particles remain together unless they actually form an effectively isolated self-gravitating system. Thus the property of isolated equilibrium configurations on which the conventional derivation of the virial theorem hinges is not present in a cosmic distribution of gravitating particles. (b) An expanding cosmic distribution is obviously very far from being an equilibrium configuration. Thus the basic premise of the conventional virial theorem is invalid.

We may regard the virial theorem from another point of view, however. In general, the instantaneous values of the forces and of the velocities in an isolated system need not be related. If they *are* related, it must be because the forces have not changed substantially over a period of time during which they have substantially altered the velocities. To formulate this idea mathematically, consider the nonrelativistic equation of motion for a particle in an isolated Newtonian system:

$$\mathbf{v}(t) = \int_{t_0}^{t} \mathbf{f}(t')\,dt' + \mathbf{v}(t_0), \tag{6.4}$$

where f denotes the force per unit mass. We regard $\mathbf{v}(t)$ and $\mathbf{f}(t)$ as random functions. From (6.4) we obtain

$$\langle v^2(t)\rangle = \int_{t_0}^{t} \langle \mathbf{v}(t)\cdot\mathbf{f}(t')\rangle\,dt' + \langle \mathbf{v}(t_0)\cdot\mathbf{v}(t)\rangle. \tag{6.5}$$

Suppose, to begin with, that the covariance of v and f is invariant under translation in time. Then we may set

$$\langle \mathbf{v}(t)\cdot\mathbf{f}(t')\rangle = \langle \mathbf{v}(t+t_1-t')\cdot\mathbf{f}(t_1)\rangle \tag{6.6}$$

where t_1 is an arbitrary fixed epoch. We then obtain from (6.5)

$$\begin{aligned}\langle v^2(t)\rangle &= -\langle \mathbf{x}(t_1)\cdot\mathbf{f}(t_1)\rangle \\ &\quad + \langle \mathbf{x}(t+t_1-t_0)\cdot\mathbf{f}(t_1)\rangle + \langle \mathbf{v}(t)\cdot\mathbf{v}(t_0)\rangle \\ &= -\langle \mathbf{x}(t)\cdot\mathbf{f}(t)\rangle \\ &\quad + \langle \mathbf{x}(t)\cdot\mathbf{f}(t_0)\rangle + \langle \mathbf{v}(t)\cdot\mathbf{v}(t_0)\rangle\end{aligned} \tag{6.7}$$

if we assume that the covariance of x and f is also translationally

invariant. Equation (6.7) reduces to the nonrelativistic form of the virial theorem (6.3) if the last two terms on the right side of (6.7) vanish. This requirement means that at time t a particle no longer remembers its dynamical state at time t_0.

To sum up, the classical virial theorem applies to a particle whose motion can be regarded as a *stationary* random process, provided that enough time has elapsed for statistical traces of the initial conditions to have been lost.

A somewhat more general theorem can be established. Instead of postulating that ensemble averages of the form $\langle p(t) \cdot q(t-\tau)\rangle$ are independent of t we make the weaker assumption that they may be adequately represented by the first two terms in a modified Taylor expansion about t:

$$\begin{aligned}\langle \mathbf{v}(t) \cdot \mathbf{f}(t-\tau)\rangle &\equiv F(t,\tau)\\ &= F(t_1+\tau,\tau) + (t-t_1-\tau)\,\partial F(t_1+\tau,\tau)/\partial t.\end{aligned} \tag{6.8}$$

We may choose t_1 so as to ensure that the integral of the second term on the right over the range $0 \leqq \tau \leqq t - t_0$ will vanish. Assuming as before, that all covariances connecting times t_0 and t vanish, we obtain the modified theorem

$$\langle v^2(t)\rangle + \langle \mathbf{x}(t_1) \cdot \mathbf{f}(t_1)\rangle = 0 \tag{6.9}$$

which expresses a relation between the expectation of the kinetic energy at time t and the virial at an earlier time t_1. The interval $(t - t_1)$ represents a time of relaxation for the system under consideration. The assumption that the first-order formula (6.8) is valid means that the time required for the particle-velocities to adjust to the forces is sufficiently small compared with a time during which the gross structure of the system changes appreciably. We shall call a system that satisfies this requirement *quasi-stationary*.

For a quasi-stationary isolated system of gravitating particles, the virial theorem takes the form

$$2\langle T(t)\rangle + \langle U(t-\epsilon)\rangle = 0, \tag{6.10}$$

where $\epsilon = t - t_1$. The assumption of quasi-stationarity allows us to replace the second term on the left side of this equation by the first two terms in its Taylor expansion:

$$2\langle T(t)\rangle + \langle U(t)\rangle - \epsilon\, d\langle U(t)\rangle/dt = 0. \tag{6.11}$$

Using the energy equation,

$$\langle T(t)\rangle + \langle U(t)\rangle = E = \text{const.}, \tag{6.12}$$

to eliminate $\langle T(t)\rangle$ and solving the resulting differential equation for $\langle U(t)\rangle$, we obtain

$$\langle U(t)\rangle = 2E - \epsilon\langle U'(t_0)\rangle \exp\left[-(t-t_0)/\epsilon\right]. \tag{6.13}$$

This equation shows that the mean potential energy of a quasi-stationary system (and hence also the mean kinetic energy) will relax to a constant value in a time comparable to the previously defined relaxation time ϵ:

$$\langle U(t)\rangle \to U = \text{const.}, \qquad \langle T(t)\rangle \to T = \text{const.} \tag{6.14}$$

The preceding derivation of the virial theorem can be applied, with appropriate modifications, to a cosmic distribution of gravitating particles. In place of Equation (6.4) we have, by (4.16),

$$\mathbf{v} = -\int_{t_0}^{t} \frac{S(t')}{S(t)} \frac{\partial\phi(\mathbf{x},t')}{\partial\mathbf{x}}\, dt' + \mathbf{v}_0 \tag{6.15}$$

whence

$$\langle v^2\rangle_m = -\int_{t_0}^{t} \left\langle \mathbf{v}(\mathbf{x},t)\frac{S(t')}{S(t)} \cdot \frac{\partial\phi(\mathbf{x},t')}{\partial\mathbf{x}} \right\rangle_m dt', \tag{6.16}$$

provided that the velocities at times t and t_0 are uncorrelated. The displacement of a particle from its initial position is given in terms of its peculiar velocity by

$$\mathbf{x}(t) - \mathbf{x}_0 = \int_{t_0}^{t} \mathbf{v}(\mathbf{x},t')\frac{S(t)}{S(t')}\, dt'. \tag{6.17}$$

An argument exactly analogous to the preceding one leads to the modified virial theorem

$$2T_m + U_m(t-\epsilon) = 0. \tag{6.18}$$

By combining this relation, expanded to first order, with the energy-balance Equation (5.8) one obtains

$$dU_m/dt, dT_m/dt \propto S^{-2}e^{-t/\epsilon}. \tag{6.19}$$

As in the case of an isolated system, the kinetic and potential energies per unit mass relax to constant values in the characteristic relaxation time ϵ.

It follows from Equation (5.10) that when the virial theorem is

satisfied the mean cosmic pressure $\bar{p}$ vanishes. This condition appears to be fulfilled at the present time.

VII. **Applications to cosmogony.** It will be shown in the next section that the virial theorem may be expected to apply whenever the mean cosmic density is substantially less than atomic density. (In a pressure-free universe this condition is satisfied when $t \gg 10^3$sec.) The kinetic and potential energies per unit mass would then remain nearly constant if it were not for radiation losses. These may have caused the values of T_m and $-U_m$ associated with galaxies to have increased by a moderate factor (probably less than 10) since the time of their formation. The present values of these quantities are rather uncertain, partly because it is not known what fraction of the matter in the universe is bound in galaxies and galaxy clusters. If this fraction is of order unity, then

$$T_m \cong 10^{14}\,\text{erg/gm}, \qquad U_m \cong -2 \times 10^{14}\,\text{erg/gm}. \tag{7.1}$$

If only 1 per cent of the matter in the universe is bound in galaxies and galaxy clusters, these estimates should be reduced by a factor of 100.

The constancy of U_m implies, according to (5.5), that

$$U_m = -2\pi G\bar{\rho}\lambda^2\alpha^2 = \text{const.} \tag{7.2}$$

This equation shows that α^2, the mean square fractional density fluctuation, must increase at least as fast as S, since $\bar{\rho}$ varies like S^{-3} and λ increases no faster than S. The higher the mean density, the smaller is the amplitude of the density fluctuations needed to store a given quantity of potential energy per unit mass. At an epoch when galaxies were just beginning to separate out, the value of α was close to unity. At earlier times, when the mean density appreciably exceeded galactic densities, $\alpha \ll 1$; the density was nearly uniform. According to (7.1) and the remark immediately following that equation, the kinetic temperature of the gas probably lay in the range 10^4—$10^{6\circ}$K.

This prediction may be confronted with observation. In the range of densities under consideration, hydrogen is completely ionized at temperatures above $10^{4\circ}$K. The acceleration of electrons by positive ions gives rise to electromagnetic radiation (bremsstrahlung, or free-free emission). Some of the photons emitted by this process will be reabsorbed by the gas, others will survive to

the present day, their number density diluted and their frequency shifted toward the red by the cosmic expansion. The intensity of the resulting background radiation field is given by the formula

$$I_{\nu_0}(t_0) = c\int_0^{t_0} s^3(t)\, j(\nu, t) \exp\left\{-c\int_t^{t_0} k_\nu dt'\right\} dt \tag{7.3}$$

where

$$s(t) = S(t)/S(t_0), \qquad \nu = s(t)\,\nu_0. \tag{7.4}$$

If we assume that local thermodynamic equilibrium prevails, the emission coefficient $j(\nu, t)$ and the absorption coefficient k_ν are related by

$$j(\nu, t) = k_\nu B(\nu, T), \tag{7.5}$$

where T is the kinetic temperature of the gas and $B(\nu, T)$ is the Planck radiation function. The form of k_ν is known from atomic theory. Miss M. Kaufman [9] has calculated the intensity of the background radiation field for temperatures in the range 10^4—$10^{6\circ}$ K. The energy of the predicted radiation field is concentrated in the microwave region. Measurements of the required accuracy in this part of the spectrum present considerable technical difficulties, but Penzias and Wilson [11] have recently succeeded in measuring what appears to be a genuine background radiation field at 4080 Mc/s. The estimated brightness temperature of 3.5° K lies in the range of values predicted for this frequency by Miss Kaufman's calculations. Proposed measurements in the millimeter region should afford a more rigorous test of the theory.[2]

VIII. **The initial stages of the expansion.** We have seen that the negative energy currently stored in galaxies and other self-gravitating systems could have been stored in density fluctuations of very small amplitude in the early stages of the expansion. There remain the questions of how the internal energy came to assume a negative value in the first place and how it came to have a spectral distribution corresponding to the observed hierarchy of astronomical systems. The answers to these questions are still far from clear,

[2] An alternative explanation of the Penzias-Wilson observation was proposed by Dicke and Peebles [10], who attribute it to a primordial radiation field whose energy density initially dominated that of the matter field and which remained in equilibrium with the matter field until its temperature dropped to about $10^{11\circ}$ K.

but I shall indicate briefly the lines along which some current studies are proceeding.

From the preceding discussion of the virial theorem it is clear that changes in T_m and U_m that occur in time-intervals comparable to the velocity-autocorrelation time of a particle need not satisfy the relation $2\delta T_m + \delta U_m = 0$. Rather, such changes will satisfy the energy-conservation law

$$\delta T_m + \delta U_m = 0, \tag{8.1}$$

provided that the velocity-autocorrelation time is much less than t, the elapsed time since the beginning of the expansion. Consequently the cosmic pressure will not be precisely zero, but will be given by

$$\bar{p} = \bar{\rho}(2\delta T_m + \delta U_m)/3 = \bar{\rho}\delta T_m/3. \tag{8.2}$$

From the energy-balance Equation (5.8) it follows that

$$dE_m/dt = -H\delta T_m < 0. \tag{8.3}$$

Thus the internal energy tends to decrease even when density fluctuations are present and the virial theorem is satisfied.

For a distribution of gravitating particles, δT_m is given by

$$\delta T_m \cong Gm/l \tag{8.4}$$

where l denotes the mean interparticle distance and m denotes the mean mass of a particle. Since $l \propto S$, it follows from (8.3) that the change in E_m during a time-interval (t_1, t_2) is given by

$$\Delta E_m \Big|_{t_1}^{t_2} \cong -(Gm/l_1 - Gm/l_2). \tag{8.5}$$

Setting $m = 10^{-24}$ gm and $l = 10^{-13}$ cm (the value appropriate for nuclear density), we obtain $\Delta E_m \cong -10^{-18}$ erg/gm—too low by more than thirty powers of 10! This result throws some light on a well-known result of Lifshitz: random fluctuations in a statistically uniform distribution of gravitating particles cannot lead to the development of significant density fluctuations of galactic scale during periods comparable to the age of the universe.

In the actual universe, the dominant interparticle forces during the early stages of the expansion are electrostatic rather than gravitational, except during a very brief interval during which nuclear forces predominate. Since the electrostatic interaction between a proton and an electron is 10^{39} times stronger than the

gravitational interaction, electrostatic interactions could conceivably produce an effect of the desired magnitude. Specifically, in order for ΔE_m to assume a final value in the neighborhood of -10^{13} erg/gm, the transition from positive to negative values of E_m should occur when the mean density $\bar{\rho}$ is close to atomic density ($\cong 1\,\mathrm{gm/cm^3}$).

What determines the internal energy during the initial stages of the expansion? Certainly, different assumptions about the "initial" state of the universe will lead to very different answers. For example, in the type of universe originally postulated by Gamow, where radiation is dominant in the initial stages, everything depends on the value assigned to an adjustable parameter that determines the value of the energy density of the radiation field at the moment when it is just equal to the energy density of the matter field. The simplest assumption about initial conditions would seem to be that at the beginning of the expansion the universe was in its lowest energy state, and it is this assumption whose consequences my students and I are currently exploring.

At densities comparable to nuclear density, the universe is in a quasi-equilibrium state at zero temperature. The internal energy is positive, however, owing to the effects of degeneracy. At later times exothermic nuclear reactions, especially those leading to the formation of helium, will make further positive contributions to the internal energy. At the same time, adiabatic cooling tends to reduce the internal energy. Preliminary calculations by Mr. P. Eltgroth [13] indicate that the transition from positive to negative values of the internal energy occurs at a density approximately equal to that suggested by the preceding considerations, i.e., at a density that would enable E_m ultimately to assume a value in the neighborhood of -10^{13} erg/gm.

The spectral distribution of gravitational energy is determined by the relative rates of growth of the Fourier components into which the fluctuating density fleld $\tilde{\rho}$ can be analyzed. At any given instant the diameter of the observable universe ($\cong 3ct$) represents an approximate upper limit for the wavelengths of coherent density fluctuations. (Essentially the same criterion results from the requirement that the radius R of a coherent density fluctuation must exceed the corresponding Schwarzschild radius $2GM/c^2 = (8\pi/3)\,G\bar{\rho}R^3/c^2$.) Thus the mass of the largest coherent fluctuation that can form at a given time is given by the approximate formula

$$M = \pi \bar{\rho} D^3/6 = (27/36)\,(c^3/G)\,t \cong 3 \times 10^{38} t, \tag{8.6}$$

since the mean density of a matter-controlled universe satisfies the relation

$$6\pi G \bar{\rho} t^2 = 1. \tag{8.7}$$

We have already estimated that the bulk of the gravitational energy is deposited in density fluctuations when $\bar{\rho} \cong 1\,\mathrm{gm/cm^3}$, i.e., when $t \cong 10^3\,\mathrm{sec}$. The corresponding value of M is

$$M \cong 3 \times 10^{41}\,\mathrm{gm} = 1.5 \times 10^8 M_\odot. \tag{8.8}$$

This is a typical value for the mass of a dwarf galaxy. On observational grounds, it is by no means unlikely that such galaxies account for a large part of the mean gravitational binding energy.

IX. **Cosmology and the second law of thermodynamics.** The picture of cosmic evolution that I have just sketched has a certain paradoxical aspect. The initial state of the universe was assumed to be very simple, and hence to require a very small quantity of information for its specification. The present state of the universe is exceedingly complex, however, and hence requires a large quantity of information for its specification. But the second law of thermodynamics states that the information contained in a macroscopic description of an isolated physical system can never increase.

A closely related paradox concerns the evolution of a universe whose mean spatial curvature is positive. The assumption that the "initial" state is one of thermodynamic equilibrium at zero temperature makes sense only if the universe returns to this state at the end of each expansion-contraction cycle. But the identity of initial and final states seems to contradict the fact that in the course of the expansion an irreversible generation of entropy (loss of information) must occur.

Finally, a problem closely related to these two paradoxes is that of accounting for the apparent anisotropy between the two directions of time. Most authors agree that this anisotropy is cosmological rather than local in origin, but there is little agreement about the precise nature of the connection between cosmology and the arrow of time.

Before embarking on a discussion of these matters, let me recall

the definitions of information and entropy, and explain what significance I attach to these quantities in the present context. Let $\{s_i\}$ denote a complete set of mutually orthogonal (i.e., mutually exclusive) states of a given physical system. To each state we assign a probability p_i. The entropy (or uncertainty) of this probability distribution is

$$H = -\sum p_i \ln p_i \geqq 0 \tag{9.1}$$

where the summation extends over the complete set of states. A diminution in the uncertainty of a stochastic description may be viewed as a gain in information about the system. Information is accordingly defined as negative entropy, apart from an additive constant that may conveniently be chosen so as to make the information vanish when the entropy assumes its maximum value subject to stated auxiliary conditions. These definitions agree with the conventional qualitative usage of the terms "uncertainty" and "information."

In the present discussion probabilities are to be regarded as physical variables rather than as numbers associated in some way with degrees of subjective uncertainty. The ergodic property of the cosmic distributions that we have been considering ensures that all probabilities are in fact measurable with arbitrary precision. Thus information and entropy are on the same footing as thermodynamic variables like pressure and density.

The entropy associated with a stochastic description of a given system depends on the *level* of the description. Suppose that the states s_i have substates s_{ij}, to which we assign probabilities p_{ij}. Let $p_{j|i}$ denote the conditional probability that a system known to be in the ith state will be found to be in the jth substate of this state. From the definition (9.1) and the relation $p_{ij} = p_{j|i}p_i$ it follows that

$$H\{p_{ij}\} = H\{p_i\} + \sum_i p_i H\{p_{j|i}\}. \tag{9.2}$$

Thus the transition from a "macroscopic" description, in which the probability distribution $\{p_i\}$ is specified, to a "microscopic" description, in which the probability distribution $\{p_{ij}\}$ is specified, results in an entropy increase represented by the second term on the right side of (9.2). This term represents the entropy specifically associated with microscopic information about the system.

The descriptions "microscopic" and "macroscopic" are relative

to one another. Thus the states of a classical N-particle system may be described with decreasing degrees of completeness by an N-particle distribution function, an $(N-1)$-particle distribution function, and so on, down to a one-particle distribution function. Each level of description is macroscopic relative to the preceding levels in this series and microscopic relative to the succeeding levels; and a description in terms of thermodynamic and hydrodynamic variables is macroscopic relative to a description in terms of the one-particle distribution function. For every pair of levels of description an equation of the form of (9.2) can be written down.

It is an immediate consequence of Liouville's theorem that the entropy associated with a *complete* statistical description of an isolated N-particle system (i.e., a description in terms of the N-particle distribution function) is constant in time. On the other hand, in certain circumstances that have not yet been defined with complete precision, the entropy associated with a stochastic description in terms of a one-particle distribution function must increase monotonically with time until thermodynamic equilibrium prevails. According to (9.2), the diminution of macroscopic information is exactly compensated by an increase of microscopic information; information flows from the macroscopic to the microscopic degrees of freedom.

Broadly speaking, a necessary condition for the law of increasing entropy to be valid is that the entropy associated with the microscopic degrees of freedom of a system should initially have its maximum possible value. That is, microscopic information about the system should initially be absent. Whenever a thermodynamic description is applicable, this condition is automatically fulfilled. In kinetic theory the situation is less clear, but it appears that in order to derive an equation for the one-particle distribution function whose solution satisfies the law of increasing entropy (Boltzmann's H theorem), one must introduce some form of the assumption that microscopic information about the initial state is absent.

The phenomenon of irreversibility in isolated physical systems has its origin in the absence of microscopic information about initial states. The assumption that *initial* states have this property singles out a direction in time. For if microscopic information about

a system is not required for the prediction of future macrostates, then it must in general be required for the prediction of past macrostates.

At first sight it seems difficult to attach objective significance to the absence of microscopic information. To begin with, as we have already noted, the dividing line between the microscopic and macroscopic domains is largely conventional. Moreover, even if we do not happen to have microscopic information about a particular physical system, we can in principle acquire it by expending a sufficient quantity of negative entropy; and a system about which we had complete microscopic information would not exhibit the arrow of time.

If the anisotropy of time is to be attributed in some way to an initial deficiency of microscopic information in the universe, the notion of microscopic information must be made more definite; and the absence of microscopic information must be shown to be an objective property of the universe.

Consider first the question of how the absence of microscopic information can be considered to be an objective property of the universe. We know that for an ordinary physical system microscopic information may be uninteresting or hard to get or both; but it can never be said to be objectively absent. On the other hand, consider a simple unbounded system, namely, a statistically uniform (Poisson) distribution of points on a line. This distribution is characterized by a single parameter, λ, the mean spacing of adjacent points. This parameter can be measured with arbitrary accuracy, given a realization of the distribution. The question arises, whether there exist two or more distinguishable realizations of the same Poisson distribution. If distinguishable realizations do exist, a given realization contains microscopic information, namely, the information that distinguishes it from its fellows; if only a single realization exists, it is *completely* specified by the value of the macroscopic parameter λ, and hence contains no microscopic information. In fact, one can easily show that any two realizations of the same distribution can be matched arbitrarily closely over any finite interval, no matter how large. Hence any two realizations are indistinguishable.

Essentially the same considerations apply to a spatially infinite

universe that is statistically homogeneous and isotropic and in which the local irregularities have a finite scale. As long as we adhere strictly to the postulate that no statistical property of the distribution serves to define a preferred position or direction in space, there is no way of distinguishing between "different" realizations of the same statistical description. The case of a finite universe raises special problems; I shall return to it presently.

We are now in a position to answer the question: what constitutes microscopic information? According to the evolutionary scheme that I outlined earlier, a finite number of data about the initial state of the universe suffices for a statistical description of its entire future history. And it follows from the preceding discussion that such a description is complete; no statement about the microscopic state of the universe could be added that would not conflict with the assumptions already made. The assumptions concerning the initial state, together with the laws of physics and their logical consequences, constitute macroscopic information, and this is the only kind of information that exists in the present description. Of course, macroscopic information in this sense may in particular circumstances include information about the states of atomic and subatomic particles.

We can now understand the anisotropy of time. The future is, by definition, the direction in which prediction is possible. According to the present theory, a complete mathematical description of the universe must unfold from a description of the initial state. It is not possible to reconstruct the past history of the universe by working backward from a complete macroscopic description of the present state.

The arrow of time is transferred in an obvious way from the universe as a whole to the astronomical systems that separate out in the course of the expansion. Every newly formed system, no matter how complex its structure may be, is devoid of microscopic information in the sense defined earlier. It is true that microscopic information about any given finite system can always be ascertained in principle; but this requires the expenditure of an equivalent amount of negative entropy. That is, the information needs to be imported into the system.

If the future is characterized by its predictability, the past is

characterized by the persistence of its traces. These objective properties of the universe—or rather of a certain kind of mathematical description of the universe—are closely analogous to properties of our subjective impression of time.

We are now in a position to resolve a paradox that I mentioned at the beginning of this discussion: How can we reconcile the second law of thermodynamics with the notion of a universe that starts from (and perhaps ends in) its simplest possible state? The justification for applying the second law of thermodynamics to the universe as a whole is the belief that entropy is an additive quantity. Thus if the law applies locally, it must also apply globally. But is entropy in fact an additive quantity over arbitrarily large volumes of space? Up to a point, the quantity of information required to describe the contents of a given volume of space undoubtedly increases in direct proportion to the volume. However, the contents of a region whose dimensions greatly exceed the scale of local irregularities is largely predictable—at least in principle—, and the accuracy of the prediction increases with the size of the region. Since only a finite quantity of information is required to specify the universe completely, the entropy per unit volume approaches zero as the volume increases indefinitely!

The preceding considerations need to be modified before they can be applied to a spatially finite universe, since in this case there obviously exist distinct realizations of the same statistical description. Because a finite universe returns to its initial state after a finite period of time, all of its possible realizations would in fact be realized in an infinite time interval. On the other hand, because each cycle begins and ends with the same state, a given realization cannot meaningfully be said to precede or follow another realization. We may therefore think of the set of realizations as a Gibbs ensemble rather than as a sequence in time. The fact that we can acquire information only about a particular member of the ensemble (the one we inhabit) is exactly analogous to the fact that in a spatially infinite universe we can in principle acquire information only about a certain finite region (the observable universe). Thus the preceding considerations of entropy and the arrow of time apply equally to finite and infinite universes satisfying the postulates of statistical homogeneity and isotropy.

References

1. E. Hubble, *The realm of the nebulae,* Yale Univ. Press, New Haven, Conn., 1936; Dover, New York, 1958.

2. G. K. Batchelor, *The theory of homogeneous turbulence,* Cambridge Univ. Press, New York, 1953.

3. A. Einstein, *The meaning of relativity,* Princeton Univ. Press, Princeton, N. J., 1950.

4. L. D. Landau and E. M. Lifshitz, *The classical theory of fields,* Pergamon Press, New York, 1962.

5. K. Lanczos, Z. Physik 31 (1925), 112.

6. A. Einstein, *Die Grundlage der allgemeinen Relativitatstheorie,* Ann. Physik 49 (1916), 109-164; transl., *The principle of relativity,* Dover, New York, 1923.

7. W. M. Irvine, *Local irregularities in an expanding universe,* Ann. Physics 32 (1965), 322-347.

8. D. Layzer, *A preface to cosmogony*: I. *The virial theorem and the energy equation for cosmic distributions,* Astrophys. J. 138 (1963), 735-737.

9. M. Kaufman, *Limits on the density of intergalactic ionized hydrogen,* Nature 207 (1965), 736-737.

10. R. H. Dicke, P. J. E. Peebles, P. G. Roll and D. T. Wilkinson, *Cosmic black body radiation,* Astrophys. J. 142 (1965), 414-419.

11. A. A. Penzias and R. W. Wilson, *A measurement of excess antenna temperature at* 4080 Mc/s, Astrophys. J. 142 (1965), 419-421.

12. D. Layzer, *The formation of stars and galaxies*: *unified hypotheses,* Ann. Rev. Astronom. Astrophys. 2 (1964), 341.

13. P. Eltgroth, *Fluctuations and correlations in the expanding universe,* Ph.D. Thesis, Harvard University, Cambridge, Massachusetts, 1966.

14. D. Layzer, *The strong cosmological principle, indeterminacy, and the direction of time,* Cornell Conference, 1963.

HARVARD COLLEGE OBSERVATORY
CAMBRIDGE, MASSACHUSETTS

W. B. Bonnor

Jeans' Formula for Gravitational Instability

1. **Introduction.** Jeans' criterion (1929) states that if p, ρ be the pressure and density of an ideal gas connected by a relation of the form $p = p(\rho)$, then a cloud of gas with linear dimension

$$l > \left(\frac{\pi}{G\rho}\frac{dp}{d\rho}\right)^{1/2} \tag{1.1}$$

where G is the gravitational constant, will be unstable to slight fluctuations in the density. It is widely used in astrophysics, and can be extended to quite general situations, for example those in which rotation or a magnetic field is present (Chandrasekhar [4]).

When the formula is derived, it is commonly assumed that the cloud is (i) static and (ii) uniform in p and ρ. However, these assumptions are unsatisfactory if gravitation is taken into account: because if the cloud is taken as finite, no solution of the hydrostatic equation $\mathbf{F} = (1/\rho)\,\nabla p$ exists; and if it is taken as infinite $\mathbf{F}$ is not defined.

In this lecture I shall derive Jeans' formula by submitting Newtonian cosmological models to small perturbations. The motive for doing this is partly to put the formula on a more secure footing, but principally to study the important problem of the formation of

galaxies in an initially uniform universe.

The pioneering work in this field was that of Lifshitz [11], who considered the stability of general relativistic cosmological models to small perturbations of a very general kind. I gave a Newtonian treatment [3] for a more restricted class of perturbations. Since that time the problem has been studied by Van Albada [17], [18], Nariai and Ueno [13], and Irvine [7]. These papers apply, in the main, only to infinitesimal perturbations. I treated the problem of condensations without ignoring nonlinear effects for a particular relativistic model [2]. The conclusion of all these works is consistent with the following statement. Expanding world models do exhibit instability to certain types of perturbation, and something like Jeans' formula is true. However, using reasonable assumptions about the magnitude of the initial perturbation, one cannot account for the formation of galaxies in those models which have zero cosmological constant.

2. **Newtonian world models.** I shall make the following assumptions:

(i) the model consists of a very large, but finite distribution* of perfect fluid with equation of state $p = p(\rho)$;

(ii) one observer O sees every particle of the model move radially with velocity u;

(iii) O sees a spherically symmetric distribution, that is, he finds p, ρ and u functions of r and t only.

The hydrodynamic equations for a perfect compressible fluid, namely

$$\partial \mathbf{u}/\partial t + (\mathbf{u} \cdot \nabla)\,\mathbf{u} = \mathbf{F} - (1/\rho)\,\nabla p, \tag{2.1}$$

$$\partial \rho/\partial t + \nabla \cdot (\rho \mathbf{u}) = 0, \tag{2.2}$$

where $\mathbf{F}$ is the body force per unit mass, yield under assumptions (i), (ii) and (iii),

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial r} = F - \frac{1}{\rho}\frac{\partial p}{\partial r}, \tag{2.3}$$

$$\frac{\partial \rho}{\partial t} + u\frac{\partial \rho}{\partial r} + \rho\left(\frac{\partial u}{\partial r} + \frac{2u}{r}\right) = 0, \tag{2.4}$$

*This assumption is made to overcome difficulties with the integrals for the gravitational field strength. See Layzer [9] and McCrea [12].

F being now the radial gravitational force per unit mass, given by

$$\nabla \cdot \mathbf{F} \equiv \partial F/\partial r + 2F/r = -4\pi G\rho + \Lambda, \tag{2.5}$$

Λ being the cosmological constant. (For justification of this equation, and for a more detailed treatment of Newtonian models, see Bondi [1].)

So far no assumption of homogeneity has been made. If we now make this by putting

$$\rho = \rho(t), \tag{2.6}$$

we can integrate (2.5) to get

$$F = -((4\pi/3)\, G\rho - (1/3)\,\Lambda)\, r, \tag{2.7}$$

ignoring a term which is singular at $r = 0$. We make also another common assumption of Newtonian cosmology, namely

$$p = 0. \tag{2.8}$$

Equations (2.3) and (2.4) now reduce to

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial r} = -\frac{4\pi}{3} G\rho r + \frac{1}{3}\Lambda r, \tag{2.9}$$

$$\frac{\partial \rho}{\partial t} + \rho\frac{\partial u}{\partial r} + \frac{2u\rho}{r} = 0, \tag{2.10}$$

where use has been made of (2.7). We now assume (cf., Bondi [1]) a solution of the form

$$u = rf(t), \tag{2.11}$$

where f is a function to be determined. Substituting (2.11) into (2.9) and (2.10) we find

$$\dot{f} + f^2 = -(4\pi/3)\, G\rho + (1/3)\,\Lambda, \tag{2.12}$$

$$\dot{\rho} + 3f\rho = 0, \tag{2.13}$$

where $\cdot$ means $\partial/\partial t$.

We pause to show that the solution (2.11) satisfies what is called the *cosmological principle*. This states that every cosmic observer (i.e., one moving with the substratum) sees, at a given instant, the same picture of the universe. Let O, O' be two cosmic observers and P an element of the substratum. Let vectors OP, $O'P$, OO'

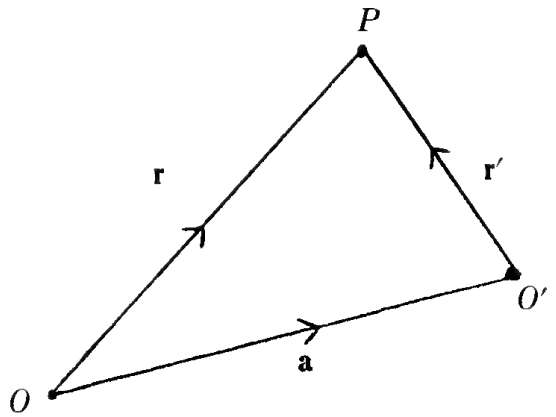


FIGURE 1

be denoted by $\mathbf{r}$, $\mathbf{r}'$, $\mathbf{a}$ as in Figure 1. Then in accordance with (2.11) O finds

$$\mathbf{u}_P = \mathbf{r}f(t) \quad \text{and} \quad \mathbf{u}'_O = \mathbf{a}f(t).$$

Subtracting we find

$$\mathbf{u}_P - \mathbf{u}_{O'} = (\mathbf{r} - \mathbf{a})\, f(t) \quad \text{or} \quad \mathbf{u}'_P = \mathbf{r}'f(t) \tag{2.14}$$

where $\mathbf{u}'_P$ means the velocity of P measured by O'. Equation (2.14) shows that every cosmic observer verifies (2.11) if O does. Since by assumption p and ρ are independent of the position of the observer, the cosmological principle is verified in the model.

We return to (2.12) and (2.13) and introduce a function $R(t)$ by

$$f = \dot{R}/R. \tag{2.15}$$

We find

$$\rho = \alpha R^{-3} \tag{2.16}$$

and

$$R^2\ddot{R} + (4\pi/3)\, G\alpha = (1/3)\, \Lambda R^3. \tag{2.17}$$

Equations (2.11), (2.16), *and* (2.17) *are the equations determining homogeneous, isotropic Newtonian world models.*

By illustration, let us put $\Lambda = 0$; then (2.17) can be integrated in terms of elementary functions. A first integral is

$$\dot{R}^2 = 8\pi G\alpha/3R - \beta^2, \tag{2.18}$$

where β is a constant of integration which can be (i) zero, (ii) imaginary, (iii) real. These three cases yield three well-known models whose counterparts in general relativity are the Einstein-deSitter, hyperbolic and cycloidal models, respectively. The graphs of the expansion functions $R(t)$ are shown in Figure 2.

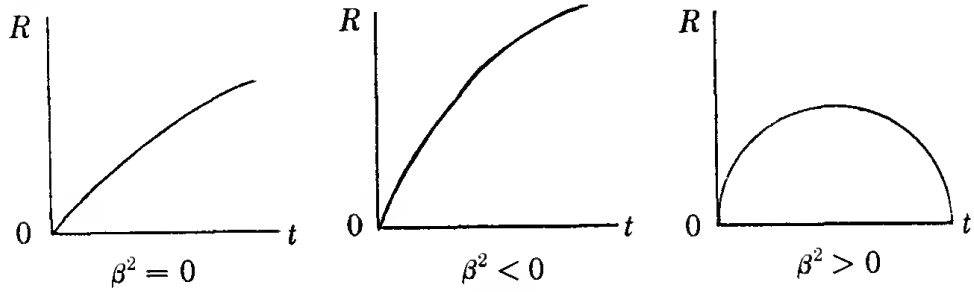


FIGURE 2

3. **Perturbation theory.** I continue to make the assumptions (i)—(iii) at the beginning of §2, though it is not difficult to do without the assumption of spherical symmetry (Savedoff and Vila [15]). The model of §2 is perturbed by putting

$$\begin{aligned} u &= r\dot{R}/R + v(r,t), \\ \rho &= \alpha R^{-3} + \omega(r,t), \\ p &= g(r,t), \end{aligned} \tag{3.1}$$

R being a given function of time, α a constant, and v, ω, g small functions whose squares and products may be neglected, and also squares and products of their derivatives. Since the pressure vanishes in the unperturbed model, we may write

$$g = \omega\, dp/d\rho. \tag{3.2}$$

We now substitute (3.1) and (3.2) into (2.3), (2.4), and (2.5). We find (Bonnor [3]) that, after a very long calculation, there emerges a single equation for ω:

$$\begin{aligned} &\ddot{\omega} + 2rf\dot{\omega}' + r^2f^2\omega'' + 8f\dot{\omega} + r\omega'(9f^2 + \dot{f}) \\ &\qquad + \omega(18f^2 + 6\dot{f} - \Lambda) = (\omega'' + 2\omega' r^{-1})\, dp/d\rho, \end{aligned} \tag{3.3}$$

where $'$ means $\partial/\partial r$ and f is given by (2.15). This equation governs the entire development of the perturbation. Equation (3.3) can be thrown into a more convenient form if we introduce a new independent variable x by

(3.4) $$r = xR(t)$$

(x *is a comoving coordinate* relative to the unperturbed model, as is verified from

$$Dx \equiv \partial x/\partial t + u \cdot \nabla x = 0),$$

and a new dependent variable s by

(3.5) $$s = \omega/\rho_0,$$

where ρ_0 $(= \alpha R^{-3})$ is the unperturbed density. The function s is the *magnitude of condensation* at any point. Equation (3.3) becomes

(3.6) $$\frac{\partial^2 s}{\partial x^2} + \frac{2}{x}\frac{\partial s}{\partial x} = \frac{R^2}{c^2}\left[\ddot{s} + \frac{2\dot{R}}{R}\dot{s} + s\left(\frac{3\ddot{R}}{R} - \Lambda\right)\right],$$

where we have introduced the velocity of sound c (in the absence of gravitation) by

(3.7) $$c^2 = dp/d\rho.$$

(3.6) is the equation of the propagation of infinitesimal, spherically symmetric sound waves in a Newtonian world model. We now consider standing waves of the form

(3.8) $$s = (\sin kx/kx)\, h(t),$$

where k is a constant. Near $x = 0$, $s \sim h(t)$, so we shall take h as a measure of the condensation. Inserting (3.8) into (3.6) we find, using (2.17), an ordinary differential equation for h:

(3.9) $$\ddot{h} + 2(\dot{R}/R)\,\dot{h} + h(k^2c^2/R^2 - 4\pi G\rho_0) = 0.$$

It is easy to see from (3.9) that condensations can form. Suppose for example that at time $t = t_1$

(3.10) $$h > 0, \qquad \dot{h} > 0;$$

the perturbation is growing, and growth continues until h reaches a maximum, i.e., until

(3.11) $$\dot{h} = 0, \qquad \ddot{h} < 0.$$

Let us now inspect (3.9) and remember that in an expanding world model $\dot{R}/R > O$; then we see that (3.11) cannot happen if

(3.12) $$c^2k^2/R^2 - 4\pi G\rho_0 < 0.$$

If (3.12) *is satisfied, perturbations initially satisfying* (3.10) *continue to grow indefinitely.*

We therefore have (3.12) as a sufficient condition for instability. Let us introduce the wavelength of the disturbance by

$$\lambda = 2\pi R/k; \tag{3.13}$$

then (3.12) becomes

$$\lambda^2 > \pi c^2/G\rho_0. \tag{3.14}$$

This is precisely equivalent to (1.1), so *we conclude that Jeans' formula is true in the expanding Newtonian world models, provided we interpret the quantities involved in an appropriate way.* In this interpretation λ, c, and ρ_0 are to be taken as time-dependent in accordance with their definitions (3.13), (3.7), and (2.16). For a consideration whether Jeans' formula applies to the steady-state model, see Bonnor [3].

4. **Formation of condensations.** Most cosmological models start from a singular state of infinite density. This is taken to represent a state of the universe in which the conditions were so severe that all matter was homogeneously distributed, except for very slight irregularities. The question then is: how has the manifest heterogeneity of the universe occurred? Or, how have the galaxies formed?

To answer this question, we have to consider carefully whether the process described in §3 could have produced condensations in the time available from the singular state to the present—about 10^{10} years in most models. In studying this we neglect all nonlinear effects. Such information as we have (Bonnor [2]) indicates that these do not affect the situation much.

I shall choose a definite Newtonian model, the cycloidal one, obtained by integrating (2.18) with β real. There is reason to suppose that this model is more favourable to condensation than the other two models with $\Lambda = 0$. The scale function R is given by

$$R = (8\pi G\alpha/3\beta^2)\sin^2(\theta/2), \tag{4.1}$$

$$t = (4\pi G\alpha/3\beta^3)(\theta - \sin\theta), \tag{4.2}$$

ignoring one uninteresting constant of integration.

Let us substitute (4.1) and (4.2) into the Equation (3.9) for the condensation growth function h. We find that (3.9) can be

expressed in the form

$$\frac{d^2h}{d\Theta^2} + \cot\left(\frac{\Theta}{2}\right)\frac{dh}{d\Theta} - \frac{3}{2}h\,\mathrm{cosec}^2\left(\frac{\Theta}{2}\right)\left(1 - \frac{\pi c^2}{G\rho_0\lambda^2}\right) = 0. \tag{4.3}$$

We now consider the most favourable case, in which $p = 0$, so $c = 0$, and the model is filled with dust. Then (4.3) can be integrated to give

$$\begin{aligned} h = &-\frac{3}{2}A\,\Theta\,\mathrm{cosec}^2\left(\frac{\Theta}{2}\right)\cot\left(\frac{\Theta}{2}\right) + A\left(2 + 3\cot^2\left(\frac{\Theta}{2}\right)\right) \\ &+ B\,\mathrm{cosec}^2\left(\frac{\Theta}{2}\right)\cot\left(\frac{\Theta}{2}\right), \end{aligned} \tag{4.4}$$

where A and B are arbitrary constants.

We turn now to the question how quickly a perturbation can grow. Suppose that it occurs at a time t_1 so early that we may approximate:

$$\cos\Theta_1 = 1 - \Theta_1^2/2, \qquad \sin\Theta_1 = \Theta_1.$$

Then (4.4) gives for the leading terms in the initial perturbation h_1

$$h_1 \sim 8B\Theta_1^{-3} + A\Theta_1^2/10. \tag{4.5}$$

We now assume that h_1 is small: then B must be very small because Θ_1 is small. However, A need not be as small as h_1.

At the present epoch $\Theta \sim 1$, so inserting this value of Θ into (4.4) and neglecting B, we have

$$\begin{aligned} &h_{\text{now}} \sim A, \quad \text{or, using (4.5),} \\ &h_{\text{now}} \sim 10\,h_1\Theta_1^{-2}. \end{aligned} \tag{4.6}$$

This gives the strength of the condensation now in terms of the initial perturbation and the time at which it occurred (Θ is connected with the time by (4.2)).

To form a numerical estimate, suppose that the perturbation occurred 1000 years after the singular state. Then we find $\Theta_1 \sim 10^{-2}$, so from (4.6)

$$h_{\text{now}} \sim 10^5 h_1. \tag{4.7}$$

Thus the initial perturbation has grown considerably. The difficulty arises when we consider the likely magnitude of h_1. If we consider h_1 to arise from fluctuations in an ideal gas we have

$$h_1 \sim \delta\rho/\rho = N^{-1/2}$$

for a collection of N molecules. For a galaxy $N \sim 10^{68}$, so $h_1 \sim 10^{-34}$, and inserting this into (4.7) we find

$$h_{\text{now}} \sim 10^{-29},$$

so that the degree of condensation is quite negligible in the case of an object of galactic size.

This conclusion is not altered if instead of galaxies we consider stars ($N \sim 10^{58}$). Nor is it altered in a meaningful way if we consider times earlier than 1000 years after the singular state. For instance, if we take the fluctuation to occur after 1 second, then $\Theta_1 \sim 10^{-6}$ and

$$h_{\text{now}} \sim 10^{13} h_1;$$

the degree of condensation is still negligible.

5. **Possible ways out.** One model in which condensation takes place quickly is the static Einstein universe. In this case R is constant, and if we take $p = 0$, the Equation (3.9) for the condensation function R becomes

$$\ddot{h} = 4\pi G\rho_0 h, \tag{5.1}$$

where ρ_0 is now a constant because this model is static. It follows that perturbations can grow like

$$\exp\left[(4\pi G\rho_0)^{1/2}t\right]. \tag{5.2}$$

If we take for illustration $\rho_0 = 10^{-23}\text{gm.cm}^{-3}$, we find that in 10^9 years the expression (5.2) increases by a factor of about 10^{36} so that galaxy formation might be possible even starting from the very small perturbations mentioned in the previous section.

This leads us to expect that galaxy formation may be possible in any model with a quasi-static period of about $10^9 - 10^{10}$ years. Such a model is that of Lemaître, which requires a positive Λ and which, in its general relativistic form, is spatially closed. The expansion function R for this model is shown in Figure 3, the quasi-static period corresponding to AB. I now think it likely that galaxies can form in Lemaître's model, and I should like to take this opportunity to retract the contrary view which I have expressed previously. Nevertheless, this conjecture needs to be con-

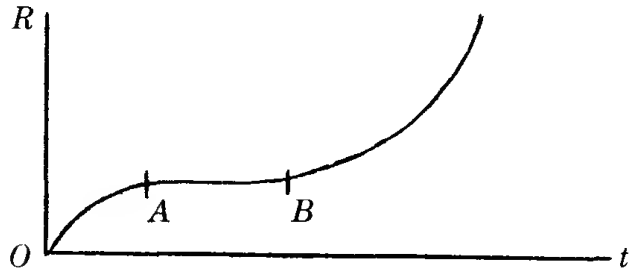


FIGURE 3

firmed by detailed calculation, using not only the Newtonian but also the relativistic equations. We also have to bear in mind the possibility that the Lemaître model is already ruled out by the red-shift-apparent magnitude observations.

Another theory in which galaxy formation can probably be explained is the steady-state model. Here the problem is quite different because large inhomogeneities are always present (Sciama [16]; Harwit [6]; Roxburgh and Saffman [14]). However, it seems that the steady-state model conflicts with observation.

A solution of a quite different kind is to keep the relativistic models with zero Λ, and to find a source of larger fluctuations. From (4.7) it follows that an initial fluctuation of order, say, 10^{-3} could lead to an incipient galaxy in the time available (assuming nonlinear effects do not impede the condensation process). It was suggested by Gamow [5] that the primordial cosmic medium was turbulent and this turbulence was the source of large fluctuations needed for galaxy formation. Leaving aside the unanswered question of the origin of the turbulence, there remains the further question whether it could have persisted long enough to assist in condensation. Layzer [10] has examined this latter question and has given a negative answer.

6. **Conclusion.** 1. We found in §3 that Jeans' formula applies to Newtonian world models provided the quantities involved are appropriately interpreted.

2. It seems that, given only the small fluctuations of statistical mechanics, the galaxies cannot form by gravitational instability in those models which have $\Lambda = 0$. Models more favourable to galaxy formation, such as Lemaître's model and the steady-state model, seem to conflict with astronomical observations.

3. If no plausible source of larger fluctuations can be found, it seems that one ought to conclude that gravitational instability has not played a major role in the formation of galaxies.

References

1. H. Bondi, *Cosmology*, Cambridge Univ. Press, New York, 1960.
2. W. B. Bonnor, Z. Astrophys. **39** (1956), 143.
3. ———, Monthly Notices Roy. Astronom. Soc. **117** (1957), 104.
4. S. Chandrasekhar, *Hydrodynamic and hydromagnetic stability*, Clarendon Press, Oxford, 1961.
5. G. Gamow, Phys. Rev. **86** (1952), 251.
6. M. Harwit, Monthly Notices Roy. Astronom. Soc. **122** (1961), 47.
7. W. M. Irvine, Ann. Phys. **32** (1965), 322.
8. J. Jeans, *Astronomy and cosmology*, Cambridge Univ. Press, New York, 1929, p.345.
9. D. Layzer, Astronom. J. **59** (1954), 268.
10. ———, Ann. Rev. Astronom. Astrophys. **2** (1964), 341.
11. E. Lifshitz, J. Phys. U.S.S.R. **10** (1946), 116.
12. W. H. McCrea, Nature **175** (1955), 466.
13. H. Nariai and Y. Ueno, Progr. Theoret. Phys. **23** (1960), 305.
14. Roxburgh and Saffman, Monthly Notices Roy. Astronom. Soc. **129** (1965), 181.
15. M. P. Savedoff and S. Vila, Astrophys. J. **136** (1962), 609.
16 D. W. Sciama, Monthly Notices Roy. Astronom. Soc. **115** (1955), 3.
17. G. B. Van Albada, Bull. Astronom. Inst. Netherlands **15** (1960), 165.
18. ———, Astronom. J. **66** (1961), 590.

QUEEN ELIZABETH COLLEGE
LONDON, ENGLAND

P. J. E. Peebles

Microwave Radiation from the Big Bang*

The search [1] for a primordial fireball is based on this idea, that the universe may have expanded from a state so dense that matter and radiation were able to strike a thermal equilibrium. This would have filled space with thermal black-body radiation, the primordial fireball.

The expansion of the universe would have cooled the fireball, but it is important that the thermal character of the radiation would have been preserved. This effect is simply described using the expansion parameter $a(t)$. As the universe expands the scale of distances in the universe varies as the expansion parameter $a(t)$, so that the mean number density of particles varies as

$$n(t) \propto a(t)^{-3}. \tag{1}$$

Neglecting for the moment any interaction between matter and radiation, photons would be conserved and the number density of photons also would satisfy (1). At the same time the photons suffer the cosmological red-shift, where the wavelength of a photon

*This research was supported in part by the National Science Foundation and the Office of Naval Research of the United States Navy.

is defined by local observers, at rest relative to the universe about them. In a time dt a photon moves between two such local observers separated by the distance cdt. Since lengths are scaling like $a(t)$ the two observers are moving apart with velocity $cdt\dot{a}/a$, so that the photon would have been red-shifted by the amount

$$d\lambda/\lambda = dt\dot{a}/a. \tag{2}$$

This yields the result [2]

$$\lambda \propto a. \tag{3}$$

For thermal radiation the number density of photons with wavelength in the range $d\lambda$ about the wavelength λ is

$$dN = n_\lambda d\lambda = 8\pi d\lambda/\lambda^4(e^{hc\ kT\lambda} - 1). \tag{4}$$

When the universe expands by a factor a the new wavelength of these photons is, from Equation (3),

$$\lambda' = \lambda a, \tag{5}$$

and from Equation (1) the new number density is

$$dN' = dN/a^3. \tag{6}$$

This means that the new photon distribution law is

$$\begin{aligned} n'_\lambda d\lambda &= \frac{1}{a^3} n_{\lambda,a} d(\lambda/a) \\ &= 8\pi\, d\lambda/\lambda^4(e^{hca\ kT\lambda} - 1). \end{aligned} \tag{7}$$

That is, the new distribution is thermal, with a new temperature

$$T' = T/a. \tag{8}$$

By similar arguments one can see that the temperature of a nonrelativistic gas would vary as

$$T_2 \propto 1/a^2 \tag{9}$$

This is simply the adiabatic cooling law for a gas uniformly distributed throughout the universe so that the gas expands as the universe expands. On the other hand if the gas were gravitationally bound in a cloud the expansion of the universe would have no effect on it.

It is seen from these results that the expansion of the universe might be traced back to a highly contracted, high temperature state. It is in this assumed state of high temperature and density that thermal relaxation would have produced a primordial fireball.

To search for the cooled fireball one would observe at wavelengths in the range of 1 to 10 cm. At wavelengths much longer than 10 cm any thermal radiation would be masked by the background due to the Galaxy and the extra-Galactic radio sources. At wavelengths much shorter than 1 cm atmospheric emission would be troublesome, and in any case the high-frequency Wien cut-off would severely restrict the intensity for any reasonable fireball temperature.

A microwave radiometer designed to search for the primordial fireball radiation at 3 cm wavelength has been built at Princeton University by P. G. Roll and D. T. Wilkinson. Before the instrument was completed a useful measure of the background came from work at the Bell Telephone Laboratories [3]. This group was using a microwave radiometer equipped with a horn-type antenna and a low-noise amplifier. In attempting to account for all sources of noise in their receiver they ended with an unexplained residual of 3.5 °K at 7.3 cm wavelength. We have suggested [1] that this residual may be the cooled fireball radiation from the big bang.

To entertain this idea we must establish first that this background could not be due to any reasonably possible source in the present state of the universe. The following is a list of possible sources of microwave background.

(1) Atmospheric emission is about 3 °K at 3 cm. It can be eliminated by observing that the emission varies as the path length through the atmosphere, that is as the secant of the angle from the vertical. The radiation intensity is measured as a function of angle from the vertical and the isotropic background extracted by curve fitting.

(2) Stellar emission is unimportant because the stars in our Galaxy subtend a very small total solid angle in the sky, amounting to perhaps 1 part in 10^{10}. At centimeter wavelengths the effective temperature of the sun can approach 10^5 °K. If all stars radiated at this rate, the integrated background would amount to $10^{-10} \times 10^5$ °K $= 10^{-5}$ °K.

(3) Interstellar dust is thought to have a temperature of about 25 °K, and from starlight extinction we know that the geometrical solid angle subtended by the dust is appreciable. However, the dust is of micron size, and at centimeter wavelengths the effective cross section for radiating is down from the geometrical by a factor at least equal to the fourth power of the ratio of wavelength to particle size, that is, roughly 10^{16}. For this reason interstellar dust is unimportant at centimeter wavelengths. Interstellar rocks could radiate effectively. However, assuming a cosmic abundance the interstellar mass density in rocks could not reasonably exceed .01 protons/cm^3, which is a spatial density less than 10^{-26}/cm^3 rocks one centimeter in radius. Since the rocks would be at a temperature of about 3 °K, looking toward the galactic pole through perhaps 1 kiloparsec of rocks the effective temperature would be of the order of $10^{-26} \times 3 \times 10^{21} \times 3 \times 3\,^\circ\mathrm{K} = 3 \times 10^{-4}\,^\circ\mathrm{K}$.

(4) Some cosmologies require that the mean mass density in the universe be about 2×10^{-29}gm/cm^3. This is a factor of 30 greater than what seems to be available in galaxies, [4] and it has been suggested that the rest is smoothly distributed intergalactic hydrogen, not yet condensed into galaxies. If this hydrogen were ionized it would radiate due to bremsstrahlung emission by the electrons accelerated in the field of the protons. The effective radiation temperature due to this process operating in the present state of the universe would amount to

$$10^{-23}\lambda^2 n^2 c/H T_P^{1/2} \lesssim 10^{-5}\,^\circ\mathrm{K} \tag{10}$$

at $\lambda = 7$ cm. Here n is the number density of intergalactic hydrogen, and T_P is the temperature, assumed not less than 10^4 °K, the minimum required to ionize the hydrogen due to collisions. The effective path length through the plasma is taken to be c/H, the dimension of the visible universe. This is the appropriate length in the steady state cosmology, or in an evolutionary cosmology in which the plasma has been turned on at a red-shift of the order of unity. In an evolutionary cosmology another possibility is that the plasma turned on in the early, more highly compressed universe. If the plasma were turned on at a time corresponding to a red-shift z, and then allowed to cool according to Equation (9), the contribution (10) would be increased by a factor $z^{3/2}$. This means that to

bring the background (10) up to 1 °K the plasma would have had to have been turned on at a red-shift $z \sim 10^3$, when the mean density in the universe was 10^9 times what it is now. This is not an appealing picture.

(5) The total radio background, galactic and extragalactic, has been measured at wavelengths longer than 75 cm. The solid line in Figure 1 is the background seen by a low resolution antenna looking out the galactic pole [5]. The straight-line power law spectrum is observed all the way back to 100 m wavelength. At wavelengths shorter than 75 cm the spectra of localized sources have been observed using higher resolution antennas, and these typically exhibit this same power law spectrum. The galactic emission observed along the galactic plane has a slope (temperature spectral index) equal to -2.7. Thermal radiation from ionized hydrogen also would have a power law spectrum, with slope -2.0. The extragalactic sources also typically exhibit an approximately straight spectrum. For most sources the observed intensity at 3 cm wavelength does not exceed the intensity extrapolated from 75 cm using a straight power law spectrum with slope -2.4 [6]. Therefore, it is reasonable to extrapolate the radio background to centimeter wavelengths according to the dotted straight line in Figure 1.

There are two sources discovered by Dent and Haddock [7], which do exhibit a sharp increase in intensity above the power law extrapolation near 3 cm wavelength. It is not likely that these sources could have caused the background of 3.5 °K at 7 cm. These peculiar spectra are a small minority among those observed so far, and furthermore in these peculiar sources the deviation from a power law spectrum is a factor of 10, while we need a factor of 100 to bring the radio source background up from the dotted extrapolation to 3.5 °K.

(6) Starlight from very distant galaxies reaches us highly red-shifted, so that the light would be displaced toward radio wavelengths. Figure 1 shows the expected integrated background for a zero pressure cosmological model with acceleration parameter $q = \frac{1}{2}$. It is assumed that the light from each galaxy is like that of a black body at 6000 °K, and that each galaxy has had constant luminosity since it turned on, when the mean density in the universe was 10^6 times what it is now. This causes a slight kink in the curve

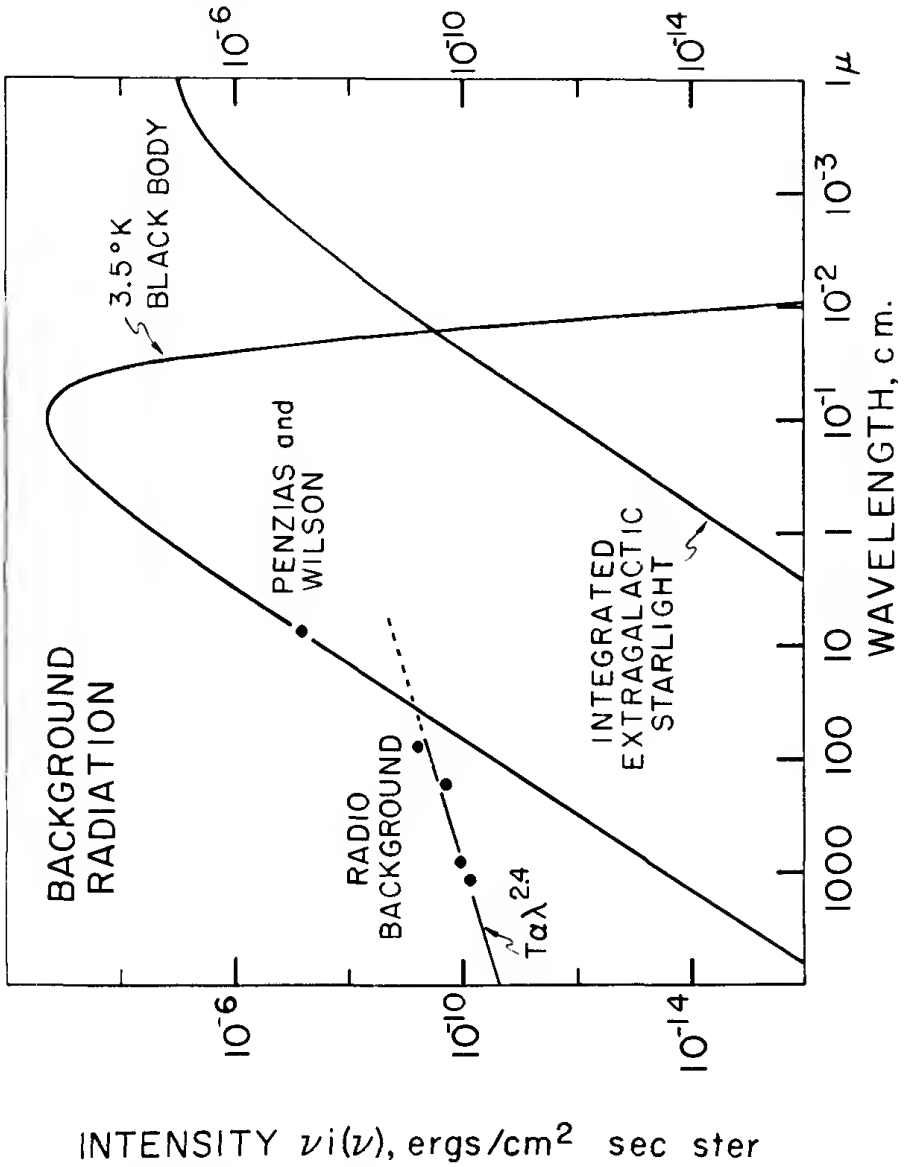


FIGURE 1. The estimated background radiation is shown for the integrated starlight from galaxies, the possible primordial fireball radiation at a temperature of 3.5 °K, and the integrated background due to the radio sources, Galactic and extra-Galactic.

at 50 μ. The calculation is uncertain on all counts, but the discrepancy is many orders of magnitude, so that no reasonable change in the model could make this an important source of background.

Apparently, no known process operating in the present state of the universe could have caused a background as high as 3.5 °K at 7 cm. This is in direct disagreement with the steady state cosmology in its original form, for here one cannot fall back on processes operating in a more highly contracted phase of the universe.

If this radiation is the primordial fireball, it should have the black-body spectrum shown in the figure. There are two possible tests for this. At wavelengths longer than 3 cm the intensity should vary as frequency squared. At shorter wavelengths the Wien cutoff should lead to an appreciable deviation from the power law spectrum, the deviation amounting to 20% at 1 cm wavelength. If these two criteria can be established it will be strong evidence that this is a black-body background, for it would be very unlikely that any other effect could fortuitously satisfy both tests.

These tests are not yet established. Nevertheless it is interesting to consider how thermal radiation of this amount could have originated—what must have been the state of the universe when the thermal radiation we observe was emitted? The answer depends on the cosmology, and we adopt as most reasonable the picture given in Table 1.

TABLE 1. Assumed Evolution of the Universe

Nucleon density (protons/cm^3)	Time	Temperature (°K)	Red-shift	
10^{-5}	$7 \times 10^9 y$	3.5	0	Now
10^4	$2 \times 10^5 y$	4000	10^3	Hydrogen Recombines
3×10^{23}	1 sec	10^{10}	3×10^9	Thermal electron pairs created

Assuming the uniform, isotropic expansion of the universe can be traced back to a red-shift of 10^3, we would reach a temperature

so high that hydrogen would have been ionized. At this epoch the mean nucleon density is 10^3 times the mean density within galaxies now. We shall suppose therefore that at this time matter was uniformly distributed.

Another interesting landmark appears if we can trace the expansion back to a red-shift $z = 3 \times 10^9$, which leads to a temperature of 10^{10}°K according to Equation (8). At this temperature the electrons are relativistic, and pair production processes raise the equilibrium electron density to the order of

$$n_e \sim aT^3/k \sim 10^{32}/\text{cm}^3, \tag{11}$$

where a is the radiation energy density constant and $T = 10^{10}$°K. With this electron density the mean free path of a photon between Thompson scatterings is

$$1/\sigma n_e \sim 10^{-7}\,\text{cm},$$

or

$$1/\sigma n_e c \sim 10^{-17}\,\text{sec}. \tag{12}$$

Since thermal energies are comparable with the electron rest mass each collision can transfer appreciable energy between electron and photon, so that this is a characteristic thermal relaxation time, 10^{17} times shorter than the characteristic expansion time of the universe (Table 1). Thus, it is seen that if the universe started expanding uniformly from this highly contracted state, then matter and radiation should have been in thermal equilibrium.

Given this initial thermal equilibrium state, it is important to ask whether the radiation would have remained thermal as the universe expanded. This is not immediately obvious because the matter tends to cool faster than the radiation, but it is easily seen that there is sufficient coupling between radiation and matter to preserve thermal equilibrium.

Apparently the most important process is free electron scattering. The magnitude of this effect may be estimated in the following way. Suppose that the free electrons have a thermal velocity distribution, at temperature T. Then a photon with energy well less than kT will gain energy, on the average, in each collision, $\langle \delta\nu \rangle / \nu = 4\langle v^2/c^2 \rangle / 3 = 3kT/mc^2$. This means that if matter and radiation are very roughly in thermal equilibrium the maximum

rate of energy transfer between matter and radiation is of the order of

$$\dot{E}/E_\gamma \cong 4\sigma nckT/mc^2,$$

where n is the electron density and E_γ is the radiation energy density. In evaluating this it is convenient to multiply by the age of the universe,

$$\begin{aligned} t\dot{E}/E_\gamma &\cong 4\sigma nct\,kT/mc^2 \\ &\cong 1.2 \times 10^{-10} z^{5/2}. \end{aligned} \tag{13}$$

The correction for stimulated emission has been neglected. In the second line z is the red-shift which a photon would suffer in propagating from the given epoch to the present, and the present mass density is assumed to be $2 \times 10^{-29}\,\mathrm{gm/cm^3}$.

The radiation energy density greatly exceeds the heat content of the plasma,

$$E_\gamma/E_m = aT^4/3nkT = 6 \times 10^7, \tag{14}$$

nearly independent of time. To preserve thermal equilibrium in the expanding universe the rate of heat transfer to the plasma must be

$$t\dot{E}/E_\gamma = \frac{2}{3}\, E_m/E_\gamma = 1 \times 10^{-8}. \tag{15}$$

When the hydrogen is thermally ionized, $z \geqq 10^3$, the maximum value (13) is more than 10^5 times larger than (15). Evidently, matter and radiation should remain at very nearly the same temperature, so that the radiation should preserve a thermal spectrum.

When the hydrogen has recombined it is adiabatically cooled faster than the radiation, but this is unimportant because the two are no longer coupled. It is apparent from (14) that the final recombination radiation from the plasma is a very small perturbation to the radiation energy density. When the galaxies form, any remaining intergalactic hydrogen may become ionized again, but emission from this hydrogen could not be appreciable at wavelengths near the peak of the fireball radiation.

If the universe is homogeneous and isotropic the primordial fireball should have a thermal spectrum near the intensity peak.

It is interesting to turn this around, and ask what limits we can place on the uniformity of the universe if the observations establish that the radiation is isotropic, with a thermal spectrum.

An isotropy measurement carries information from the region where the radiation last was scattered. If there is ionized intergalactic hydrogen amounting to a mean density of 2×10^{-29} gm/cm^3, the optical depth for scattering by the free electrons is unity at a red-shift $z = 7$, when the age of the universe is $t = 3 \times 10^8$ years. The electron scattering would smooth the radiation over a characteristic distance comparable with the size ct of the visible universe at this time. Taking account of expansion this amounts to smoothing over one third of the size of the present visible universe. An isotropy observation thus provides information on the gross structure of the visible universe, but if the early universe were irregular over a smaller scale than this, the irregularities in the background radiation might be smoothed over by scattering off intergalactic electrons.

The spectrum of the radiation depends on the structure of the universe down to a smaller scale. If the temperature in the early universe were irregular over a scale of distances somewhat smaller than the smoothing provided by electron scattering, the radiation seen now would consist of a superposition of black-body spectra, rather than a true thermal spectrum. This is because the photon energy is changed only slightly in each collision with a free electron. Going back to a red-shift in excess of $z = 1 \times 10^4$ it is seen from Equation (13) that at these early times collisions are so frequent that there is appreciable energy transfer due to free electron scattering. This means that the radiation would have been able to redistribute its energy over a thermal spectrum. This minimum red-shift for thermal relaxation corresponds to an age of the universe $t = 6 \times 10^3$ years, and the mass within the visible universe at this time is $10^{16} M_\odot$.

Apparently, if it exists the primordial fireball provides a remarkably powerful test of the uniformity of the early universe. The isotropy and spectrum of the radiation depends on the structure of the universe, on a scale extending from the universe visible at present, and going down to distances encompassing $10^{16} M_\odot$ when the age of the universe was 6×10^3 years.

References

1. R. H. Dicke, P. J. E. Peebles, P. G. Roll and D. T. Wilkinson, Astrophys. J. 142 (1965), 414.

2. C. Callan, R. H. Dicke and P. J. E. Peebles, Amer. J. Phys. 33 (1965), 105.

3. A. A. Penzias and R. W. Wilson, Astrophys. J. 142 (1965), 419.

4. J. H. Oort, Solvay Conf. on *Structure and evolution of the universe*, R. Stoops ed., Coudenberg, Brussels, Belgium, 1958, p. 163.

5. A. J. Turtle, J. F. Pugh, S. Kenderdine and I. I. K. Pauliny-Toth, Monthly Notices Roy. Astronom. Soc. 124 (1962), 297.

6. W. A. Dent and F. T. Haddock, *Quasi-stellar sources and gravitational collapse*, ed. I. Robinson, A. Schild and E. L. Schücking, University of Chicago Press, Chicago, 1965, p. 381.

7. ———, Nature 205 (1965), 487.

PRINCETON UNIVERSITY
PRINCETON, NEW JERSEY

Index

absolute derivative, 24
action, 43
advance of perihelion of orbit of inner planet, 107
algebraically
  general, 97, 102
  special, 97, 102
    metrics, 160
asymptotic flatness, 152
asymptotically flat space-times, 148
atlas, 62
atomic hydrogen in intergalactic medium, neutral, 207

back-scattering of gravitational waves, 149
bending of light, 92
Bianchi identity, 25
bicharacteristics, 67
Birkhoff, see ergodic theorems
bivectors, 96
Bondi-Sachs "news function", 159

Cartan frames, 201
Cauchy problem, 61, 68
center of mass
  local, 18
  relativistic, 18
characteristic(s), 67
  hypersurfaces, 66
charge, 26
  total, 14
Christoffel symbol, 24
classical tests, 84
clustering
  amplitude, 247
  scale, 247
coefficients of connection, 46
collapse
  with pressure-gradient, 198
  zero-pressure, 196
collision laws, 26
co-moving, 185
  coordinates, 187
    stationary in, 189
condensation(s), 268, 269, 271
conformal
  technique, 150
  tensor, 101, 154
  transformation, 150
conservation
  of charge, 13
  laws, 48, 50
  of mass, 173
  of momentum, 177

conserved quantities, 148
constraint equations, 70
continuum mechanics, 15
coordinate
  invariance, 106
  patch, 162
cosmic evolution, 256
cosmogony, 251
cosmological
  constant, 43
  principle, 265, 266
cosmology, 237
  relativistic, 229
covariant derivative, 24
cross section, 145
  absorption, 144
curvature, 242
  radius of, 242
  tensor, 141

Debever-Penrose directions, 102
density
  discontinuity, 183
  fluctuation, mean square fractional, 252
  scalar rest mass, 172
differentiable manifold, 62
differential forms, 165
directions, principal null, 97, 102
distribution function(s), 170, 258
dual, 12, 24
duality rotation, 97
dust, 16, 232

Eddington, 143
Einstein, 142, 143
  -deSitter cycloidal model, 267, 269
  -deSitter hyperbolic model, 267
  tensor, 25
  three tests, 107
  universe, static, 271
elastic solid, normal modes of, 144
electromagnetic potentials, 13, 26
electromagnetism, 12
  Maxwell's equations, 26
energy
  binding, 196
  equation, 246
  -momentum tensor, 232

entropy, 257
  irreversible generation of, 256
  S, specific, 176
equivalence, principle of, 20
ergodic
  property, 257
  theorems of Birkhoff and Khinchin, 239
ergodicity, 246
Eulerian
  coordinate system, 187
  velocities, 185
expansion parameter $a(t)$, 274

Fermi
  coordinates,
    of first kind, 22
    geodesic, 23, 55
    of second kind, 54
  derivative, 53
  propagation, 53
field equations
  gravitational, 37
  weak, 200
fluctuations, 241
fluids, perfect, 175
free particle, 26
Friedman, 239
  equation, 225
  models, 224, 228
fundamental form
  first, 72
  second, 73

galaxy
  cluster, 238
  formation, 271, 272
galaxies, 264, 272
  formation of, 264, 269, 273
gauge transformations, 135
Gauss' divergence theorem, 157
Gaussian normal coordinates, 71
geodesic
  affine parameter, 24
  coordinates, 21
  deviation, 40, 41
  Fermi coordinates, 23, 55
  variational principle, 25
gravitation constant, 143

gravitational, 211
  collapse, 194
  deflection of star light, 107
  field equations, 37
  plane waves, 130
  rotation, 141, 148
  wave antenna, 144
  wave detector, 145
gravitons, 142
Green's Theorem, 13
gyroscope, 105
  precession, 108

harmonic oscillator, 144
heat flow-vector, 174
Hertz type experiments, 143, 146
hierarchy of self-gravitating systems, 238
Hugoniot, see Rankine
hypersurface(s), null, 65, 149
hydrostatic pressure, 174

inertial
  frames, 200
    dragging of, 200
  system, 221
information, 256, 257
initial
  stages of expansion, 253
  value equation, 201
interplanetary plasma, 126
isolated system, 14, 17

Jeans' formula, 263, 264, 269
jump conditions, 61
junction conditions, 63

Khinchin, see ergodic theorems
Killing vector field, 48
Killing's equation, 48
Kruskal extension, 81

Lagrange coordinates, 185, 187
Lagrangian, 43, 185
laws, local, 219
Lemaître model, 271, 272
Levi-Civita symbol, 10

Lie
  derivative, 47
  differential, 47
limiting fluid, 178
Lorentz
  force, 13
  equations of motion, 26
  metric, 142
  transformations, 1, 4
Lyman α line, 207

Mach, 245
Mach's principle, 201
macroscopic, 257
magnitude of condensation, 268
manifolds, 161
mass, 148
  center, 17
  -current effect, 110
material stress energy tensor, 15, 27
matter, incoherent, 16
Maxwell stress energy tensor, 16, 26
metric(s)
  analytic continuation for, 161
  tensor, 106
microscopic, 257
  information, 260
Minkowski space-time, 1
momentum, 148
  angular, 17
    density, 17
  conservation of, 177
  world, 11, 17, 26
motion, equations of, 51

neutron star, 211
Newman-Penrose gravitational invariants, 148
news function, 159
Newtonian
  approximation, 38
  cosmological models, 263
  world models, 264, 265, 266, 268, 269, 272
"nonrotating" coordinate systems, 245
null, 97
  cone, 2

NUT space, see Taub

orthonormal tetrad, 22

parallel propagation, 24
particle,
  pole-dipole, 56
  simple pole, 56
  test, 55
partition, 98, 102
"peeling-off" property, 157
Penrose
  see Debever, Newman
Petrov-Pirani classification, 94, 102
photon, 26
piezoelectric
  crystals, 145
  effect, 146
planetary radar reflection, 112
polarization, 133
precession
  geodetic, 110
  Thomas, 109
primordial fireball, 274

quadrupole moments
  rotational, 138
  vibrational, 138
quasi-stationarity, 250
quasi-stationary isolated systems of gravitating particles, 250
quasi-stellar radio sources, 203, 204

radar
  astronomy, 117
  performance, 117
radiation, background, 278
random process, stationary, 250
Range-Doppler mapping, 119
Rankine-Hugoniot equations, 178
red shift, 85, 204, 211
  acceleration, 34
  cosmological, 274
  gravitational, 30, 107
  magnitude relation, 207
  second order, 110
relaxation time, 145
rest specific
  enthalpy, 175
  internal energy, 174
Riemann
  normal coordinate system, 144
  tensor
    discontinuity of vacuum, 65
    measurement of, 143
Riemannian space, 21
Ricci tensor, curvature invariant, 25
rotating mass shell, 201

Schwarzschild
  metric, isotropic, 201
    solution, 73
    sphere, 78
shock wave(s), 183
singularity, 163
deSitter, see Einstein
slip-stream discontinuity, 183
space, affine, 46
spin
  tensor, 59
  two field, 59
  vector, 59
spinor(s), 6, 94
  curvature, 100
  two-component, 155
static, 73
steady-state
  cosmology, 269, 272, 279
stress tensor, 174
substratum, 239
  equation of state for, 240

Taub-NUT space, 161
temperature, 176
tetrad components, 95
time
  delay, 125
  dilatation
    acceleration, 34
    gravitational, 30
  proper, 3
  of relaxation, 250

thermodynamics, second law of, 256
total stress energy tensor, 16
twin effect, 3
turbulence, 272

ultraviolet excess, 205
uncertainty, 257

uniformity, principle of, 238
universe
  observable, 238
  radius of the observable, 242

variational principle, 43
vorticity vector, 191

wave
  fronts, 136
  vector, 11
Weyl tensor (conformal curvature tensor), 25
world
  force, 11, 26
  line, 2
  velocity, 10

Lectures in Applied Mathematics

Proceedings of the Summer Seminar, Boulder, Colorado, 1960

VOLUME I — LECTURES IN STATISTICAL MECHANICS
G. E. Uhlenbeck and G. W. Ford with E. W. Montroll

VOLUME II — MATHEMATICAL PROBLEMS OF RELATIVISTIC PHYSICS
I. E. Segal with G. W. Mackey

VOLUME III — PERTURBATION OF SPECTRA IN HILBERT SPACE
K. O. Friedrichs

VOLUME IV — QUANTUM MECHANICS
V. Bargmann

Proceedings of the Summer Seminar, Ithaca, New York, 1963

VOLUME V — SPACE MATHEMATICS, PART 1
J. Barkley Rosser, Editor

VOLUME VI — SPACE MATHEMATICS, PART 2
J. Barkley Rosser, Editor

VOLUME VII — SPACE MATHEMATICS, PART 3
J. Barkley Rosser, Editor

Proceedings of the Summer Seminar, Ithaca, New York, 1965

VOLUME VIII — RELATIVITY THEORY AND ASTROPHYSICS
1. RELATIVITY AND COSMOLOGY
Jürgen Ehlers, Editor

VOLUME IX — RELATIVITY THEORY AND ASTROPHYSICS
2. GALACTIC STRUCTURE
Jürgen Ehlers, Editor

VOLUME X — RELATIVITY THEORY AND ASTROPHYSICS
3. STELLAR STRUCTURE
Jürgen Ehlers, Editor